

**Original citation:**

Doan, Xuan Vinh and Vavasis, Stephen. (2016) Finding the largest low-rank clusters with Ky Fan 2-k-norm and l1-norm. SIAM Journal on Optimization, 26 (1). pp. 274-312.

**Permanent WRAP url:**

<http://wrap.warwick.ac.uk/74985>

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

"First Published in SIAM Journal on Optimization in 26 (1) published by the Society for Industrial and Applied Mathematics (SIAM). Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

**A note on versions:**

The version presented in WRAP is the published version or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: [publications@warwick.ac.uk](mailto:publications@warwick.ac.uk)

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk/>

## FINDING THE LARGEST LOW-RANK CLUSTERS WITH KY FAN 2- $k$ -NORM AND $\ell_1$ -NORM\*

XUAN VINH DOAN<sup>†</sup> AND STEPHEN VAVASIS<sup>‡</sup>

**Abstract.** We propose a convex optimization formulation with the Ky Fan 2- $k$ -norm and  $\ell_1$ -norm to find  $k$  largest approximately rank-one submatrix blocks of a given nonnegative matrix that has low-rank block diagonal structure with noise. We analyze low-rank and sparsity structures of the optimal solutions using properties of these two matrix norms. We show that, under certain hypotheses, with high probability, the approach can recover rank-one submatrix blocks even when they are corrupted with random noise and inserted into a much larger matrix with other random noise blocks.

**Key words.** low rank matrix approximation, nonnegative matrix factorization, Ky Fan norm, subgaussian random noise, biclustering

**AMS subject classifications.** 15A23, 15A60, 15B52, 90C22, 62H30

**DOI.** 10.1137/140962097

**1. Introduction.** Given a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$  that has low-rank block diagonal structure with noise, we would like to find that low-rank block structure of  $\mathbf{A}$ . Doan and Vavasis [6] have proposed a convex optimization formulation to find a large approximately rank-one submatrix of  $\mathbf{A}$  with the nuclear norm and  $\ell_1$ -norm. The proposed LAROS problem (“large approximately rank-one submatrix”) in [6] can be used to sequentially extract features in data. For example, given a corpus of documents in some language, it can be used to co-cluster (or bicluster) both terms and documents, i.e., to identify simultaneously both subsets of terms and subsets of documents strongly related to each other from the *term-document matrix*  $\mathbf{A} \in \mathbb{R}^{m \times n}$  of the underlying corpus of  $n$  documents with  $m$  defined terms (see, for example, Dhillon [4]). Here, “term” means a word in the language, excluding common words such as articles and prepositions. The  $(i, j)$  entry of  $\mathbf{A}$  is the number of occurrences of term  $i$  in document  $j$ , perhaps normalized. Another example is the biclustering of gene expression data to discover expression patterns of gene clusters with respect to different sets of experimental conditions (see the survey by Madeira and Oliveira [16] for more details). Gene expression data can be represented by a matrix  $\mathbf{A}$  whose rows are in correspondence with different genes and columns are in correspondence with different experimental conditions. The value  $a_{ij}$  is the measurement of the expression level of gene  $i$  under the experimental condition  $j$ .

If the selected terms in a bicluster occur proportionally in the selected documents, we can intuitively assign a topic to that particular term-document bicluster. Similarly, if the expression levels of selected genes are proportional in all selected experiments of a bicluster in the second example, we can identify an expression pattern for the given

\*Received by the editors March 24, 2014; accepted for publication (in revised form) November 25, 2015; published electronically January 26, 2016. This work was supported in part by the U.S. Air Force Office of Scientific Research, a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada, and a grant from MITACS.

<http://www.siam.org/journals/siopt/26-1/96209.html>

<sup>†</sup>DIMAP and ORMS Group, Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK (xuan.doan@wbs.ac.uk). This research was partly done while the author was at the Department of Combinatorics and Optimization, University of Waterloo, Canada.

<sup>‡</sup>Department of Combinatorics and Optimization, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada (vavasis@uwaterloo.ca).

gene-experimental condition bicluster. Mathematically, for each bicluster  $i$ , we obtain a subset  $\mathcal{I}_i \subset \{1, \dots, m\}$  and  $\mathcal{J}_i \subset \{1, \dots, n\}$  and the submatrix block  $\mathbf{A}(\mathcal{I}_i, \mathcal{J}_i)$  is approximately rank one, i.e.,  $\mathbf{A}(\mathcal{I}_i, \mathcal{J}_i) \approx \mathbf{w}_i \mathbf{h}_i^T$ . Assuming there are  $k$  biclusters and  $\mathcal{I}_i \cap \mathcal{I}_j = \emptyset$  and  $\mathcal{J}_i \cap \mathcal{J}_j = \emptyset$  for all  $i \neq j$ , we then have the following approximation:

$$(1.1) \quad \mathbf{A} \approx [\bar{\mathbf{w}}_1, \dots, \bar{\mathbf{w}}_k][\bar{\mathbf{h}}_1, \dots, \bar{\mathbf{h}}_k]^T,$$

where  $\bar{\mathbf{w}}_i$  and  $\bar{\mathbf{h}}_i$  are the zero-padded extensions of  $\mathbf{w}_i$  and  $\mathbf{h}_i$  to vectors of length  $m$  and  $n$ , respectively. If the matrix  $\mathbf{A}$  is nonnegative and consists of these  $k$  (row- and column-exclusive) biclusters, we may assume that  $\mathbf{w}_i, \mathbf{h}_i \geq \mathbf{0}$  for all  $i$  (a consequence of the Perron–Frobenius theorem; see, for example, Golub and Van Loan [9] for more details). Thus  $\mathbf{A} \approx \mathbf{W}\mathbf{H}^T$ , where  $\mathbf{W}, \mathbf{H} \geq \mathbf{0}$ , which is an approximate *nonnegative matrix factorization* (NMF) of the matrix  $\mathbf{A}$ . In this paper, we shall follow the NMF representation to find row- and column-exclusive biclusters. Note that there are different frameworks for biclustering problems such as the graph partitioning models used in Dhillon [4], Tanay, Sharan, and Shamir [19], and Ames [1], among other models (see, for example, the survey by Nan, Boyko, and Pardalos [7]).

Approximate and exact NMF problems are difficult to solve. The LAROS problem proposed by Doan and Vavasis [6] can be used as a subroutine for a greedy algorithm with which columns of  $\mathbf{W}$  and  $\mathbf{H}$  are constructed sequentially. Each pair of columns corresponds to a feature (or pattern) in the original data matrix  $\mathbf{A}$ . Given the properties of the LAROS problem, the most significant feature (in size and magnitude) will be constructed first with the appropriate parameter.

The iterated use of the LAROS algorithm of [6] to extract blocks one at a time, however, will not succeed in the case that there are two or more hidden blocks of roughly the same magnitude. In order to avoid this issue, we propose a new convex formulation that allows us to extract several (nonoverlapping) features simultaneously. In section 2, we study the proposed convex relaxation and the properties of its optimal solutions. In section 3, we provide conditions to recover low-rank block structure of the block diagonal data matrix  $\mathbf{A}$  in the presence of random noise. Finally, we demonstrate our results with some numerical examples in section 4, including a synthetic biclustering example and a synthetic gene expression example from the previous literature.

**Notation.**  $\langle \mathbf{A}, \mathbf{X} \rangle = \text{trace}(\mathbf{A}^T \mathbf{X})$  is used to denote the inner product of two matrices  $\mathbf{A}$  and  $\mathbf{X}$  in  $\mathbb{R}^{m \times n}$ .  $\|\mathbf{X}\|_1$  means the sum of the absolute values of all entries of  $\mathbf{X}$ , i.e., the  $\ell_1$ -norm of  $\text{vec}(\mathbf{X})$ , the long vector constructed by the concatenation of all columns of  $\mathbf{X}$ . Similarly,  $\|\mathbf{X}\|_\infty$  is the maximum absolute value of entries of  $\mathbf{X}$ , i.e., the  $\ell_\infty$ -norm of  $\text{vec}(\mathbf{X})$ .

**2. Matrix norm minimization.** We start with the following general norm minimization problem, which has been considered in [6]:

$$(2.1) \quad \begin{array}{ll} \min & \|\mathbf{X}\| \\ \text{s.t.} & \langle \mathbf{A}, \mathbf{X} \rangle \geq 1, \end{array}$$

where  $\|\cdot\|$  is an arbitrary norm function on  $\mathbb{R}^{m \times n}$ . The associated dual norm  $\|\cdot\|^\star$  is defined as

$$(2.2) \quad \begin{array}{ll} \|\mathbf{A}\|^\star = \max & \langle \mathbf{A}, \mathbf{Y} \rangle \\ \text{s.t.} & \|\mathbf{Y}\| \leq 1. \end{array}$$

These two optimization problems are closely related and their relationship is captured in the following lemmas and theorem discussed in Doan and Vavasis [6].

LEMMA 2.1. Matrix  $\mathbf{X}^*$  is an optimal solution of problem (2.1) if and only if  $\mathbf{Y}^* = (\|\mathbf{A}\|_*^*) \mathbf{X}^*$  is an optimal solution of problem (2.2).

LEMMA 2.2. The set of all optimal solutions of problem (2.2) is the subdifferential of the dual norm function  $\|\cdot\|_*$  at  $\mathbf{A}$ ,  $\partial\|\mathbf{A}\|_*^*$ .

THEOREM 2.3 (Doan and Vavasis [6]). The following statements are true:

- (i) The set of optimal solutions of problem (2.1) is  $(\|\mathbf{A}\|_*^*)^{-1} \partial\|\mathbf{A}\|_*^*$ , where  $\partial\|\cdot\|_*$  is the subdifferential of the dual norm function  $\|\cdot\|_*$ .
- (ii) Problem (2.1) has a unique optimal solution if and only if the dual norm function  $\|\cdot\|_*$  is differentiable at  $\mathbf{A}$ .

The LAROS problem in [6] belongs to a special class of (2.1) with parametric matrix norms of the form  $\|\mathbf{X}\|_\theta = \|\mathbf{X}\| + \theta\|\mathbf{X}\|_1$ , where  $\|\cdot\|$  is a unitarily invariant norm and  $\theta$  is a nonnegative parameter,  $\theta \geq 0$ :

$$(2.3) \quad \begin{aligned} \min \quad & \|\mathbf{X}\| + \theta\|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \langle \mathbf{A}, \mathbf{X} \rangle \geq 1. \end{aligned}$$

A norm  $\|\cdot\|$  is unitarily invariant if  $\|\mathbf{U}\mathbf{X}\mathbf{V}\| = \|\mathbf{X}\|$  for all pairs of unitary matrices  $\mathbf{U}$  and  $\mathbf{V}$  (see, for example, Lewis [15] for more details). For the LAROS problem,  $\|\mathbf{X}\|$  is the nuclear norm,  $\|\mathbf{X}\| = \|\mathbf{X}\|_*$ , which is the sum of singular values of  $\mathbf{X}$ . In order to characterize the optimal solutions of (2.3), we need to compute the dual norm  $\|\cdot\|_*$ :

$$(2.4) \quad \begin{aligned} \|\mathbf{A}\|_\theta^* = \max \quad & \langle \mathbf{A}, \mathbf{Y} \rangle \\ \text{s.t.} \quad & \|\mathbf{Y}\| + \theta\|\mathbf{Y}\|_1 \leq 1. \end{aligned}$$

The following proposition, which is a straightforward generalization of Proposition 7 in [6], provides a dual formulation to compute  $\|\cdot\|_\theta^*$ .

PROPOSITION 2.4. The dual norm  $\|\mathbf{A}\|_\theta^*$  with  $\theta > 0$  is the optimal value of the following optimization problem:

$$(2.5) \quad \begin{aligned} \|\mathbf{A}\|_\theta^* = \min \quad & \max \{ \|\mathbf{Y}\|_*, \theta^{-1} \|\mathbf{Z}\|_\infty \} \\ \text{s.t.} \quad & \mathbf{Y} + \mathbf{Z} = \mathbf{A}. \end{aligned}$$

The optimality conditions of (2.3) are described in the following proposition, which is again a generalization of Proposition 9 in [6].

PROPOSITION 2.5. Consider a feasible solution  $\mathbf{X}$  of problem (2.3). If there exists  $(\mathbf{Y}, \mathbf{Z})$  that satisfies the conditions

- (i)  $\mathbf{Y} + \mathbf{Z} = \mathbf{A}$  and  $\|\mathbf{Y}\|_* = \theta^{-1} \|\mathbf{Z}\|_\infty$ ,
- (ii)  $\mathbf{X} \in \alpha \partial\|\mathbf{Y}\|_*$ ,  $\alpha \geq 0$ ,
- (iii)  $\mathbf{X} \in \beta \partial\|\mathbf{Z}\|_\infty$ ,  $\beta \geq 0$ ,
- (iv)  $\alpha + \theta\beta = (\|\mathbf{A}\|_\theta^*)^{-1}$ ,

then  $\mathbf{X}$  is an optimal solution of problem (2.3). In addition, if

- (v)  $\|\cdot\|_*$  is differentiable at  $\mathbf{Y}$  or  $\|\cdot\|_\infty$  is differentiable at  $\mathbf{Z}$ ,
- then  $\mathbf{X}$  is the unique optimal solution.

The low-rank structure of solutions obtained from the LAROS problem comes from the fact that the dual norm of the nuclear norm is the spectral norm (or 2-norm),  $\|\mathbf{X}\| = \sigma_1(\mathbf{X})$ , the largest singular value of  $\mathbf{X}$ . More exactly, it is due to the structure of the subdifferential  $\partial\|\cdot\|$ . According to Ziętak [22], if  $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  is a singular value decomposition of  $\mathbf{Y}$  and  $s$  is the multiplicity of the largest singular value of  $\mathbf{Y}$ , the subdifferential  $\partial\|\mathbf{Y}\|$  is written as follows:

$$\partial\|\mathbf{Y}\| = \left\{ \mathbf{U} \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T : \mathbf{S} \in \mathcal{S}_+^s, \|\mathbf{S}\|_* = 1 \right\},$$

where  $\mathcal{S}_+^s$  is the set of positive semidefinite matrices of size  $s$ . The description of the subdifferential shows that the maximum possible rank of  $\mathbf{X} \in \alpha \partial \|\mathbf{Y}\|$  is the multiplicity of the largest singular value of  $\mathbf{Y}$  and if  $s = 1$ , we achieve rank-one solutions. This structural property of the subdifferential  $\partial \|\cdot\|$  motivates the norm optimization formulation for the LAROS problem, which aims to find a *single* approximately rank-one submatrix of the data matrix  $\mathbf{A}$ . We now propose a new pair of norms that would allow us to handle several approximately rank-one submatrices simultaneously instead of individual ones. Let us consider the following norm, which we call the Ky Fan 2- $k$ -norm given its similar formulation to that of the classical Ky Fan  $k$ -norm:

$$(2.6) \quad \|\mathbf{A}\|_{k,2} = \left( \sum_{i=1}^k \sigma_i^2(\mathbf{A}) \right)^{\frac{1}{2}},$$

where  $\sigma_1 \geq \dots \sigma_k \geq 0$  are the first  $k$  largest singular values of  $\mathbf{A}$ ,  $k \leq k_0 = \text{rank}(\mathbf{A})$ . The dual norm of the Ky Fan 2- $k$ -norm is denoted by  $\|\cdot\|_{k,2}^*$ . According to Bhatia [3], the Ky Fan 2- $k$ -norm is a  $Q$ -norm, which is unitarily invariant (Definition IV.2.9 in [3]). Since the Ky Fan 2- $k$ -norm is unitarily invariant, we can define its corresponding symmetric gauge function,  $\|\cdot\|_{k,2} : \mathbb{R}^n \rightarrow \mathbb{R}$ , as follows:

$$(2.7) \quad \|\mathbf{x}\|_{k,2} = \left( \sum_{i=1}^k |x|_{(i)}^2 \right)^{\frac{1}{2}},$$

where  $|x|_{(i)}$  is the  $(n-i+1)$ st order statistic of  $|\mathbf{x}|$ . The dual norm of this gauge function (or more exactly, its square), has been used in Argyriou, Foygel, and Srebro [2] as a regularizer in sparse prediction problems. More recently, its matrix counterpart is considered in McDonald, Pontil, and Stamos [17] as a special case of the matrix cluster norm defined in [13], whose square is used for multitask learning regularization. On the other hand, the square Ky Fan 2- $k$ -norm is considered as a penalty in low-rank regression analysis in Giraud [8]. In this paper, we are going to use the dual Ky Fan 2- $k$ -norm, not its square, in our formulation given its structural properties, which will be explained later.

When  $k = 1$ , the Ky Fan 2- $k$ -norm becomes the spectral norm, whose subdifferential has been used to characterize the low-rank structure of the optimal solutions of the LAROS problem. We now propose the following optimization problem, of which the LAROS problem is a special instance with  $k = 1$ :

$$(2.8) \quad \begin{aligned} \min \quad & \|\mathbf{X}\|_{k,2}^* + \theta \|\mathbf{X}\|_1 \\ \text{s.t.} \quad & \langle \mathbf{A}, \mathbf{X} \rangle \geq 1, \end{aligned}$$

where  $\theta$  is a nonnegative parameter,  $\theta \geq 0$ . The proposed formulation is an instance of the parametric problem (2.3) and we can use results obtained in Propositions 2.4 and 2.5 to characterize its optimal solutions. Before doing so, we first provide an equivalent semidefinite optimization formulation for (2.8) in the following proposition.

PROPOSITION 2.6. Assuming  $m \geq n$ , the optimization problem (2.8) is then equivalent to the following semidefinite optimization problem:

$$(2.9) \quad \begin{aligned} \min_{p, \mathbf{P}, \mathbf{Q}, \mathbf{R}, \mathbf{X}} \quad & p + \text{trace}(\mathbf{R}) + \theta \langle \mathbf{E} \mathbf{Q} \rangle \\ \text{s.t.} \quad & kp - \text{trace}(\mathbf{P}) = 0, \\ & p\mathbf{I} - \mathbf{P} \succeq 0, \\ & \begin{pmatrix} \mathbf{P} & -\frac{1}{2}\mathbf{X}^T \\ -\frac{1}{2}\mathbf{X} & \mathbf{R} \end{pmatrix} \succeq 0, \\ & \mathbf{Q} \geq \mathbf{X}, \mathbf{Q} \geq -\mathbf{X}, \\ & \langle \mathbf{A}, \mathbf{X} \rangle \geq 1, \end{aligned}$$

where  $\mathbf{E}$  is the matrix of all ones.

*Proof.* We first consider the dual norm  $\|\mathbf{X}\|_{k,2}^*$ . We have that

$$(2.10) \quad \begin{aligned} \|\mathbf{X}\|_{k,2}^* = \max \quad & \langle \mathbf{X} \mathbf{Y} \rangle \\ \text{s.t.} \quad & \|\mathbf{Y}\|_{k,2} \leq 1. \end{aligned}$$

Since  $m \geq n$ , we have that  $(\|\mathbf{Y}\|_{k,2})^2 = \|\mathbf{Y}^T \mathbf{Y}\|_k$ , where  $\|\cdot\|_k$  is the Ky Fan  $k$ -norm, i.e., the sum of  $k$  largest singular values. Since  $\mathbf{Y}^T \mathbf{Y}$  is symmetric,  $\|\mathbf{Y}^T \mathbf{Y}\|_k$  is actually the sum of  $k$  largest eigenvalues of  $\mathbf{Y}^T \mathbf{Y}$ . Similar to  $\|\mathbf{x}\|_k$ , which is the sum of  $k$  largest elements of  $\mathbf{x}$ , we obtain the following (dual) optimization formulation for  $\|\mathbf{Y}^T \mathbf{Y}\|_k$  (for example, see Laurent and Rendl [14]):

$$\begin{aligned} \|\mathbf{Y}^T \mathbf{Y}\|_k = \min \quad & kz + \text{trace}(\mathbf{U}) \\ \text{s.t.} \quad & z\mathbf{I} + \mathbf{U} \succeq \mathbf{Y}^T \mathbf{Y}, \\ & \mathbf{U} \succeq 0. \end{aligned}$$

Applying the Schur complement, we have that

$$\begin{aligned} \|\mathbf{Y}^T \mathbf{Y}\|_k = \min \quad & kz + \text{trace}(\mathbf{U}) \\ \text{s.t.} \quad & \begin{pmatrix} z\mathbf{I} + \mathbf{U} & \mathbf{Y}^T \\ \mathbf{Y} & \mathbf{I} \end{pmatrix} \succeq 0, \\ & \mathbf{U} \succeq 0. \end{aligned}$$

Thus, the dual norm  $\|\cdot\|_{k,2}^*$  can be computed as follows:

$$\begin{aligned} \|\mathbf{X}\|_{k,2}^* = \max \quad & \langle \mathbf{X} \mathbf{Y} \rangle \\ \text{s.t.} \quad & kz + \text{trace}(\mathbf{U}) \leq 1, \\ & \begin{pmatrix} z\mathbf{I} + \mathbf{U} & \mathbf{Y}^T \\ \mathbf{Y} & \mathbf{I} \end{pmatrix} \succeq 0, \\ & \mathbf{U} \succeq 0. \end{aligned}$$

Applying the strong duality theory under Slater's condition, we have that

$$(2.11) \quad \begin{aligned} \|\mathbf{X}\|_{k,2}^* = \min \quad & p + \text{trace}(\mathbf{R}) \\ \text{s.t.} \quad & kp - \text{trace}(\mathbf{P}) = 0, \\ & p\mathbf{I} - \mathbf{P} \succeq 0, \\ & \begin{pmatrix} \mathbf{P} & -\frac{1}{2}\mathbf{X}^T \\ -\frac{1}{2}\mathbf{X} & \mathbf{R} \end{pmatrix} \succeq 0. \end{aligned}$$

The reformulation of  $\|\mathbf{X}\|_1$  is straightforward with the new decision variable  $\mathbf{Q}$  and additional constraints  $\mathbf{Q} \geq \mathbf{X}$  and  $\mathbf{Q} \geq -\mathbf{X}$ , given the fact that the main problem is a minimization problem.  $\square$

Proposition 2.6 indicates that in general, we can solve (2.8) by solving its equivalent semidefinite optimization formulation (2.9) with any SDP solver. We are now ready to study some properties of optimal solutions of (2.8). We have that  $\|\mathbf{X}\|_{k,2}^* + \theta \|\mathbf{X}\|_1$  is a norm for  $\theta \geq 0$  and we denote it by  $\|\mathbf{X}\|_{k,2,\theta}$ . According to Proposition 2.4, the dual norm  $\|\mathbf{X}\|_{k,2,\theta}^*$ ,

$$(2.12) \quad \begin{aligned} \|\mathbf{A}\|_{k,2,\theta}^* &= \max_{\mathbf{X}} \langle \mathbf{A}, \mathbf{X} \rangle \\ \text{s.t.} \quad &\|\mathbf{X}\|_{k,2,\theta} \leq 1, \end{aligned}$$

can be calculated by solving the following optimization problem given  $\theta > 0$ :

$$(2.13) \quad \begin{aligned} \|\mathbf{A}\|_{k,2,\theta}^* &= \min_{\mathbf{Y}, \mathbf{Z}} \max \{ \|\mathbf{Y}\|_{k,2}, \theta^{-1} \|\mathbf{Z}\|_\infty \} \\ \text{s.t.} \quad &\mathbf{Y} + \mathbf{Z} = \mathbf{A}. \end{aligned}$$

Similar to Proposition 2.5, we can provide the optimality conditions for (2.8) in the following proposition.

**PROPOSITION 2.7.** *Consider a feasible solution  $\mathbf{X}$  of problem (2.8). If there exists  $(\mathbf{Y}, \mathbf{Z})$  that satisfies the conditions*

- (i)  $\mathbf{Y} + \mathbf{Z} = \mathbf{A}$  and  $\|\mathbf{Y}\|_{k,2} = \theta^{-1} \|\mathbf{Z}\|_\infty$ ,
- (ii)  $\mathbf{X} \in \alpha \partial \|\mathbf{Y}\|_{k,2}$ ,  $\alpha \geq 0$ ,
- (iii)  $\mathbf{X} \in \beta \partial \|\mathbf{Z}\|_\infty$ ,  $\beta \geq 0$ ,
- (iv)  $\alpha + \theta\beta = (\|\mathbf{A}\|_{k,2,\theta}^*)^{-1}$ ,

*then  $\mathbf{X}$  is an optimal solution of Problem (2.3). In addition, if*

- (v)  $\|\cdot\|_{k,2}$  is differentiable at  $\mathbf{Y}$  or  $\|\cdot\|_\infty$  is differentiable at  $\mathbf{Z}$ ,

*then  $\mathbf{X}$  is the unique optimal solution.*

The optimality conditions presented in Proposition 2.7 indicate that some properties of optimal solutions of (2.8) can be derived from the structure of  $\partial \|\cdot\|_{k,2}$ . We shall characterize the subdifferential  $\partial \|\cdot\|_{k,2}$  next. According to Watson [21], since  $\|\cdot\|_{k,2}$  is a unitarily invariant norm,  $\partial \|\mathbf{A}\|_{k,2}$  is related to  $\partial \|\sigma(\mathbf{A})\|_{k,2}$ , where  $\sigma(\mathbf{A})$  is the vector of singular values of  $\mathbf{A}$ . Let  $\mathbf{A} \neq \mathbf{0}$  be a matrix with singular values that satisfy

$$\sigma_1 \geq \cdots > \sigma_{k-t+1} = \cdots = \sigma_k = \cdots = \sigma_{k+s} > \cdots \geq \sigma_p,$$

where  $p = \min\{m, n\}$ , so that the multiplicity of  $\sigma_k$  is  $s + t$ . The subdifferential  $\partial \|\sigma\|_{k,2}$  is characterized in the following lemma.

**LEMMA 2.8.**  *$\mathbf{v} \in \partial \|\sigma\|_{k,2}$  if and only if  $\mathbf{v}$  satisfies the following conditions:*

- (i)  $v_i = \frac{\sigma_i}{\|\sigma\|_{k,2}}$  for all  $i = 1, \dots, k - t$ .
- (ii)  $v_i = \tau_i \frac{\sigma_k}{\|\sigma\|_{k,2}}$ ,  $0 \leq \tau_i \leq 1$ , for all  $i = k - t + 1, \dots, k + s$ , and  $\sum_{i=k-t+1}^{k+s} \tau_i = t$ .
- (iii)  $v_i = 0$  for all  $i = k + s + 1, \dots, p$ .

*Proof.* Let  $\mathcal{N}_k$  be the collection of all subsets with  $k$  elements of  $\{1, \dots, p\}$ , we have

$$\|\sigma\|_{k,2} = \max_{N \in \mathcal{N}_k} f_N(\sigma),$$

where  $f_N(\sigma) = (\sum_{i \in N} \sigma_i^2)^{\frac{1}{2}}$  for all  $N \in \mathcal{N}_k$ . According to the Dubovitski–Milyutin theorem (see, for example, Tikhomirov [20]), the subdifferential of  $\|\cdot\|_{k,2}$  is computed as follows:

$$\partial \|\sigma\|_{k,2} = \text{conv} \left\{ \partial f_N(\sigma) : N \in \mathcal{N}_k, f_N(\sigma) = \|\sigma\|_{k,2} \right\}.$$



With the structure of  $\sigma$ , clearly  $\{1, \dots, k-t\} \in N$  for all  $N \in \mathcal{N}_k$  such that  $f_N(\sigma) = \|\sigma\|_{k,2}$ . The remaining  $t$  elements of  $N$  are chosen from  $s+t$  values from  $\{k-t+1, \dots, k+s\}$ . Since  $\sigma \neq \mathbf{0}$ , all  $f_N$  that satisfy  $f_N(\sigma) = \|\sigma\|_{k,2}$  are differentiable at  $\sigma$  (even in the case  $\sigma_k = 0$ ) and

$$\frac{\partial f_N(\sigma)}{\partial \sigma_i} = \frac{\sigma_i}{\|\sigma\|_{k,2}} \quad \forall i \in N, \quad \frac{\partial f_N(\sigma)}{\partial \sigma_i} = 0, \quad i \notin N.$$

Thus if  $\mathbf{v} \in \partial \|\sigma\|_{k,2}$  for all  $i = 1, \dots, k-t$ , then,  $v_i = \frac{\sigma_i}{\|\sigma\|_{k,2}}$  and  $v_i = 0$  for all  $i = k+s+1, \dots, p$ .

Finally, counting arguments for the appearance of each index in  $\{k-t+1, \dots, k+s\}$  with respect to all subsets  $N \in \mathcal{N}_k$  that satisfy  $f_N(\sigma) = \|\sigma\|_{k,2}$  allows us to characterize  $v_i$  for  $i = k-t+1, \dots, k+s$  as  $v_i = \tau_i \frac{\sigma_i}{\|\sigma\|_{k,2}}$ ,  $0 \leq \tau_i \leq 1$ , and  $\sum_{i=k-t+1}^{k+s} \tau_i = t$ .  $\square$

We are ready to characterize the subdifferential of  $\|\cdot\|_{k,2}$  with the following proposition.

**PROPOSITION 2.9.** *Consider  $\mathbf{A} \neq \mathbf{0}$ . Let  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$  be a particular singular value decomposition of  $\mathbf{A}$  and assume that  $\sigma(\mathbf{A})$  satisfies  $\sigma_1 \geq \dots > \sigma_{k-t+1} = \dots = \sigma_k = \dots = \sigma_{k+s} > \dots \geq \sigma_p$ . Then,  $\mathbf{G} \in \partial \|\mathbf{A}\|_{k,2}$  if and only if there exists  $\mathbf{T} \in \mathbb{R}^{(s+t) \times (s+t)}$  such that*

$$\mathbf{G} = \frac{1}{\|\mathbf{A}\|_{k,2}} \left( \mathbf{U}_{[:,1:k-t]} \Sigma_{[1:k-t,1:k-t]} \mathbf{V}_{[:,1:k-t]}^T + \sigma_k \mathbf{U}_{[:,k-t+1:k+s]} \mathbf{T} \mathbf{V}_{[:,k-t+1:k+s]}^T \right),$$

where  $\mathbf{T}$  is symmetric positive semidefinite,  $\|\mathbf{T}\| \leq 1$ , and  $\|\mathbf{T}\|_* = t$ .

*Proof.* According to Watson [21], we have that

$$\partial \|\mathbf{A}\|_{k,2} = \left\{ \mathbf{U} \text{Diag}(\mathbf{g}) \mathbf{V}^T : \mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T \text{ is any SVD of } \mathbf{A}, \mathbf{g} \in \partial \|\sigma(\mathbf{A})\|_{k,2} \right\}.$$

Let  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$  be a particular singular value decomposition of  $\mathbf{A}$  and assume that a singular value  $\sigma_i > 0$  has the multiplicity of  $r$  with corresponding singular vectors  $\mathbf{U}_i \in \mathbb{R}^{m \times r}$  and  $\mathbf{V}_i \in \mathbb{R}^{n \times r}$ . Then for any singular value decomposition of  $\mathbf{A}$ ,  $\mathbf{A} = \bar{\mathbf{U}}\bar{\Sigma}\bar{\mathbf{V}}^T$ , there exists an orthonormal matrix  $\mathbf{W} \in \mathbb{R}^{r \times r}$ ,  $\mathbf{W}\mathbf{W}^T = \mathbf{I}$ , such that  $\bar{\mathbf{U}}_i = \mathbf{U}_i\mathbf{W}$  and  $\bar{\mathbf{V}}_i = \mathbf{V}_i\mathbf{W}$  (for example, see Ziętak [22]).

Combining these results with Lemma 2.8, the proof is straightforward with a singular value (or eigenvalue) decomposition of matrix  $\mathbf{T}$ .  $\square$

**COROLLARY 2.10.**  $\|\cdot\|_{k,2}$  is differentiable at any  $\mathbf{A} \neq \mathbf{0}$  such that  $\sigma_k > \sigma_{k+1}$  ( $\sigma_{p+1} = 0$ ) or  $\sigma_k = 0$ .

*Proof.* If  $\sigma_k = 0$ , then, according to Proposition 2.9,

$$\mathbf{G} \in \partial \|\mathbf{A}\|_{k,2} \Leftrightarrow \mathbf{G} = \frac{1}{\|\mathbf{A}\|_{k,2}} \mathbf{U}_{[:,1:k-t]} \Sigma_{[1:k-t,1:k-t]} \mathbf{V}_{[:,1:k-t]}^T.$$

Now, if  $\sigma_k > \sigma_{k+1}$ , then  $s = 0$ . Thus  $\mathbf{T} = \mathbf{I}$  is unique since  $\mathbf{T} \in \mathcal{S}^t$ ,  $\|\mathbf{T}\|_* = t$ , and  $\|\mathbf{T}\| \leq 1$ . Thus  $\partial \|\mathbf{A}\|_{k,2}$  is a singleton, which implies  $\|\cdot\|_{k,2}$  is differentiable at  $\mathbf{A}$ .  $\square$

Proposition 2.9 shows that problem (2.8) with  $\theta = 0$  is a convex optimization problem that finds  $k$ -approximation of a matrix  $\mathbf{A}$ . It also shows that intuitively, problem (2.8) can be used to recover  $k$  largest approximately rank-one submatrices



with  $\theta > 0$ . Note that for Ky Fan  $k$ -norm, if  $\sigma_k(\mathbf{A}) > \sigma_{k+1}(\mathbf{A})$ , its subdifferential at  $\mathbf{A}$  is a singleton with a unique subgradient:

$$\partial \|\mathbf{A}\|_k = \left\{ \mathbf{U} \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{V}^T \right\},$$

where  $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$  is a singular value decomposition of  $\mathbf{A}$  and  $\mathbf{I}_k$  is the identity matrix in  $\mathbb{R}^{k \times k}$  (see, for example, Watson [21]). In this particular case, the unique subgradient of the Ky Fan  $k$ -norm provides the information of singular vectors corresponding to the  $k$  largest singular values. Having said that, it does not preserve the information of singular values. When  $\theta = 0$ , the proposed formulation with the Ky Fan  $k$ -norm will not return the rank- $k$  approximation of the matrix  $\mathbf{A}$  as the Ky Fan 2- $k$ -norm does. In the next section, we shall study the recovery of these submatrices under the presence of random noise.

**3. Recovery with block diagonal matrices and random noise.** We consider  $\mathbf{A} = \mathbf{B} + \mathbf{R}$ , where  $\mathbf{B}$  is a block diagonal matrix, each block having rank one, while  $\mathbf{R}$  is a noise matrix. The main theorem shows that under certain assumptions concerning the noise, the positions of the blocks can be recovered from the solution of (2.8). As mentioned in the introduction, this corresponds to solving a special case of the approximate NMF problem, that is, a factorization  $\mathbf{A} \approx \mathbf{W}\mathbf{H}^T$ , where  $\mathbf{W}$  and  $\mathbf{H}$  are nonnegative matrices. The special case solved is that  $\mathbf{W}$  and  $\mathbf{H}$  each consist of nonnegative columns with nonzeros in disjoint positions (so that  $\mathbf{A}$  is approximately a matrix with disjoint blocks each of rank one). Even this special case of NMF is NP-hard unless further restrictions are placed on the data model given the fact that the (exact) LAROS problem is NP-hard (see [6] for details).

Before starting the proof of the theorem, we need to consider some properties of subgaussian random variables. A random variable  $x$  is  $b$ -subgaussian if  $\mathbb{E}[x] = 0$  and there exists a  $b > 0$  such that for all  $t \in \mathbb{R}$ ,

$$(3.1) \quad \mathbb{E} [e^{tx}] \leq e^{\frac{b^2 t^2}{2}}.$$

We can apply the Markov inequality for the  $b$ -subgaussian random variable  $x$  and obtain the following inequalities:

$$(3.2) \quad \mathbb{P}(x \geq t) \leq \exp(-t^2/(2b^2)) \text{ and } \mathbb{P}(x \leq -t) \leq \exp(-t^2/(2b^2)) \quad \forall t > 0.$$

The next three lemmas, which show several properties of random matrices and vectors with independent subgaussian entries, are adopted from Doan and Vavasis [6] and references therein.

**LEMMA 3.1.** *Let  $x_1, \dots, x_k$  be independent  $b$ -subgaussian random variables and let  $a_1, \dots, a_k$  be scalars that satisfy  $\sum_{i=1}^k a_i^2 = 1$ . Then  $x = \sum_{i=1}^k a_i x_i$  is a  $b$ -subgaussian random variable.*

**LEMMA 3.2.** *Let  $\mathbf{B} \in \mathbb{R}^{m \times n}$  be a random matrix, where  $b_{ij}$  are independent  $b$ -subgaussian random variables for all  $i = 1, \dots, m$ , and  $j = 1, \dots, n$ . Then for any  $u > 0$ ,*

$$\mathbb{P}(\|\mathbf{B}\| \geq u) \leq \exp \left( - \left( \frac{8u^2}{81b^2} - (\log 7)(m+n) \right) \right).$$

LEMMA 3.3. Let  $\mathbf{x}, \mathbf{y}$  be two vectors in  $\mathbb{R}^n$  with independent and identically distributed (i.i.d.)  $b$ -subgaussian entries. Then for any  $t > 0$ ,

$$\begin{aligned}\mathbb{P}(\mathbf{x}^T \mathbf{y} \geq t) &\leq \exp\left(-\min\left\{\frac{t^2}{(4eb^2)^2 n}, \frac{t}{4eb^2}\right\}\right) \text{ and} \\ \mathbb{P}(\mathbf{x}^T \mathbf{y} \leq -t) &\leq \exp\left(-\min\left\{\frac{t^2}{(4eb^2)^2 n}, \frac{t}{4eb^2}\right\}\right).\end{aligned}$$

With these properties of subgaussian variables presented, we are now able to state and prove the main theorem, which gives sufficient conditions for optimization problem (2.8) to recover  $k$  blocks in the presence of noise.

THEOREM 3.4. Suppose  $\mathbf{A} = \mathbf{B} + \mathbf{R}$ , where  $\mathbf{B}$  is a block diagonal matrix with  $k_0$  blocks, that is,  $\mathbf{B} = \text{diag}(\mathbf{B}_1, \dots, \mathbf{B}_{k_0})$ , where  $\mathbf{B}_i = \bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T$ ,  $\bar{\mathbf{u}}_i \in \mathbb{R}^{m_i}$ ,  $\bar{\mathbf{v}}_i \in \mathbb{R}^{n_i}$ ,  $\|\bar{\mathbf{u}}_i\|_2 = \|\bar{\mathbf{v}}_i\|_2 = 1$ ,  $\bar{\mathbf{u}}_i > \mathbf{0}$ ,  $\bar{\mathbf{v}}_i > \mathbf{0}$  for all  $i = 1, \dots, k_0$ . Assume the blocks are ordered so that  $\bar{\sigma}_1 \geq \bar{\sigma}_2 \geq \dots \geq \bar{\sigma}_{k_0} > 0$ . Matrix  $\mathbf{R}$  is a random matrix composed of blocks in which each entry is a translated  $b$ -subgaussian variable, i.e., there exists  $\mu_{ij} \geq 0$  such that elements of the matrix block  $\mathbf{R}_{ij}/(\phi_i \phi_j)^{1/2} - \mu_{ij} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T$  are independent  $b$ -subgaussian random variables for all  $i, j = 1, \dots, k_0$ . Here  $\phi_i = \bar{\sigma}_i / \sqrt{m_i n_i}$ ,  $i = 1, \dots, k_0$ , is a scaling factor to match the scale of  $\mathbf{R}_{ij}$  with that of  $\mathbf{B}_i$  and  $\mathbf{B}_j$ , and  $\mathbf{e}_m$  denotes the  $m$ -vector of all 1's.

We define the following positive scalars that control the degree of heterogeneity among the first  $k$  blocks:

$$\begin{aligned}(3.3) \quad \delta_u &\leq \min_{i=1, \dots, k} \|\bar{\mathbf{u}}_i\|_1 / \sqrt{m_i}, \\ (3.4) \quad \delta_v &\leq \min_{i=1, \dots, k} \|\bar{\mathbf{v}}_i\|_1 / \sqrt{n_i}, \\ (3.5) \quad \xi_u &\leq \min_{i=1, \dots, k} \left( \min_{j=1, \dots, m_i} \bar{u}_{i,j} \right) \sqrt{m_i}, \\ (3.6) \quad \xi_v &\leq \min_{i=1, \dots, k} \left( \min_{j=1, \dots, n_i} \bar{v}_{i,j} \right) \sqrt{n_i}, \\ (3.7) \quad \pi_u &\geq \max_{i=1, \dots, k} \left( \max_{j=1, \dots, m_i} \bar{u}_{i,j} \right) \sqrt{m_i}, \\ (3.8) \quad \pi_v &\geq \max_{i=1, \dots, k} \left( \max_{j=1, \dots, n_i} \bar{v}_{i,j} \right) \sqrt{n_i}, \\ (3.9) \quad \rho_m &\geq \max_{i,j=1, \dots, k} m_i / m_j, \\ (3.10) \quad \rho_n &\geq \max_{i,j=1, \dots, k} n_i / n_j, \\ (3.11) \quad \rho_\sigma &\geq \bar{\sigma}_1 / \bar{\sigma}_k.\end{aligned}$$

We also assume that the blocks do not diverge much from being square; more precisely we assume that  $m_i \leq O(n_j^2)$  and  $n_i \leq O(m_j^2)$  for  $i, j = 1, \dots, k$ . Let  $\mathbf{p} = (k, \delta_u, \delta_v, \pi_u, \pi_v, \xi_u, \xi_v, \rho_\sigma, \rho_m, \rho_n)$  denote the vector of parameters controlling the heterogeneity.

For the remaining noise blocks  $i = k+1, \dots, k_0$ , we assume that their dominant singular values are substantially smaller than those of the first  $k$ ,

$$(3.12) \quad \bar{\sigma}_{k+1} \leq \frac{0.23 \bar{\sigma}_k}{k+1},$$

that their scale is bounded,

$$(3.13) \quad \phi_i \leq c_0 \phi_j$$

for all  $i = k+1, \dots, k_0$  and  $j = 1, \dots, k$ , where  $c_0$  is a constant, and that their size is bounded,

$$(3.14) \quad \sum_{i=k+1}^{k_0} (m_i + n_i) \leq c_1(\mathbf{p}, c_0, b) \min_{i=1, \dots, k} m_i n_i,$$

where  $c_1(\mathbf{p}, c_0, b)$  is given by (3.123) below. Assume that

$$(3.15) \quad \mu_{ij} \leq c_2(\mathbf{p}, c_0)$$

for all  $i, j = 1, \dots, k_0$ , where  $c_2(\mathbf{p}, c_0)$  is given by (3.118) below. Then provided that

$$(3.16) \quad c_3(\mathbf{p}) \left( \sum_{i=1}^k m_i n_i \right)^{-1/2} \leq \theta \leq 2c_3(\mathbf{p}) \left( \sum_{i=1}^k m_i n_i \right)^{-1/2},$$

where  $c_3(\mathbf{p})$  is given by (3.115) below, the optimization problem (2.8) will return  $\mathbf{X}$  with nonzero entries precisely in the positions of  $\mathbf{B}_1, \dots, \mathbf{B}_k$  with probability exponentially close to 1 as  $m_i, n_i \rightarrow \infty$  for all  $i = 1, \dots, k_0$ .

*Remarks.*

1. Note that the theorem does not recover the exact values of  $(\bar{\sigma}_i, \bar{\mathbf{u}}_i, \bar{\mathbf{v}}_i)$ ; it is clear that this is impossible in general under the assumptions made.
2. The theorem is valid under arbitrary permutation of the rows and columns (i.e., the block structure may be “concealed”) since (2.8) is invariant under such transformations.
3. Given the fact that for all  $i = 1, \dots, k$ ,

$$0 < \left( \min_{j=1, \dots, m_i} \bar{u}_{i,j} \right) \sqrt{m_i} \leq \|\bar{\mathbf{u}}_i\|_1 / \sqrt{m_i} \leq 1 \leq \left( \max_{j=1, \dots, m_i} \bar{u}_{i,j} \right) \sqrt{m_i},$$

we can always choose  $\xi_u$ ,  $\delta_u$ , and  $\pi_u$  such that  $0 < \xi_u \leq \delta_u \leq 1 \leq \pi_u$ . Similarly, we assume  $0 < \xi_v \leq \delta_v \leq 1 \leq \pi_v$ . These parameters measure how much  $\bar{\mathbf{u}}_i$  and  $\bar{\mathbf{v}}_i$  diverge from  $\mathbf{e}_{m_i}$  and  $\mathbf{e}_{n_i}$  after normalization, respectively. The best case for our theory (i.e., the least restrictive values of parameters) occurs when all of these scalars are equal to 1. Similarly  $\rho_\sigma, \rho_m, \rho_n \geq 1$ , and the best case for the theory is when they are all equal to 1.

4. It is an implicit assumption of the theorem that the scalars contained in  $\mathbf{p}$  as well as  $b$ , which controls the subgaussian random variables, stay fixed as  $m_i, n_i \rightarrow \infty$ .
5. As compared to the recovery result in Ames [1] for the planted  $k$ -biclique problem, our result for the general bicluster problem is in general weaker in terms of noise magnitude (as compared to data magnitude) but stronger in terms of block sizes. Ames [1] requires  $m_i = \tau_i^2 n_i$ , where  $\tau_i$  are scalars for all  $i$ ,  $i = 1, \dots, k+1$ , whereas we only need  $m_i \leq O(n_j^2)$  and  $n_i \leq O(m_j^2)$  for  $i, j = 1, \dots, k$ . More importantly, the noise block size,  $n_{k+1}$ , is more restricted as compared to data block sizes,  $n_i$ , for  $i = 1, \dots, k$ , in Ames [1] with the condition

$$c_1 \left( \sqrt{k} + \sqrt{n_{k+1}} + 1 \right) \sqrt{\sum_{i=1}^{k+1} n_i + \beta \tau_{k+1} n_{k+1}} \leq c_2 \gamma \min_{i=1, \dots, k} n_i.$$

In contrast, for our recovery result, (3.14) means that the total size of the noise blocks can be much larger (approximately the square) than the size of the data blocks. Thus, the theorem shows that the  $k$  blocks can be found even though they are hidden in a much larger matrix. In the special case when  $k_0 = k + 1$ ,  $m_i = n_i = n$ ,  $\bar{\sigma}_i = \bar{\sigma}$  for all  $i = 1, \dots, k$ , and  $m_{k+1} = n_{k+1}$ , combining (3.13) and (3.14), we will obtain the following condition, which clearly shows the relationship between block sizes:

$$\frac{\bar{\sigma}_{k+1}}{c_0 \bar{\sigma}} n \leq n_{k+1} \leq \frac{c_1(\mathbf{p}, c_0, b)}{2} n^2.$$

6. As compared to the recovery result in Doan and Vavasis [6] when  $k = 1$ , our recovery result is for a more general setting with  $\bar{\sigma}_2 > 0$  instead of  $\bar{\sigma}_2 = 0$  as in Doan and Vavasis [6]. We therefore need additional conditions on  $\bar{\sigma}_i$ ,  $i = 1, 2$ . In addition, we need to consider the off-diagonal blocks  $(i, j)$  for  $i, j = 1, \dots, k$ , which is not needed when  $k = 1$ . This leads to more (stringent) conditions on the noise magnitudes. Having said that, the conditions on the parameter  $\theta$  and block sizes remain similar. We still require  $\theta$  to be in the order of  $(m_1 n_1)^{-1/2}$  as in Doan and Vavasis [6]. The conditions  $m_1 \leq O(n_1^2)$  and  $n_1 \leq O(m_1^2)$  are similar to the condition  $m_1 n_1 \geq \Omega((m_1 + n_1)^{4/3})$  in Doan and Vavasis [6]. Finally, the condition  $m_2 + n_2 \leq c_1(\mathbf{p}, c_0, b) m_1 n_1$  is close to the condition  $m_1 n_1 \geq \Omega(m_1 + m_2 + n_1 + n_2)$ , which again shows the similarity of these recovery results in terms of block sizes.

In order to simplify the proof, we first consolidate all blocks  $i = k+1, \dots, k_0$  into a single block and call it block  $(k+1)$  of size  $\bar{m}_{k+1} \times \bar{n}_{k+1}$  where  $\bar{m}_{k+1} = \sum_{i=k+1}^{k_0} m_i$  and  $\bar{n}_{k+1} = \sum_{i=k+1}^{k_0} n_i$ . The only difference is that the new block  $\bar{\mathbf{B}}_{k+1, k+1} \in \mathbb{R}^{\bar{m}_{k+1} \times \bar{n}_{k+1}}$  is now a block diagonal matrix with  $k_0 - k$  blocks instead of a rank-one block. Similarly, new blocks  $\bar{\mathbf{R}}_{i, k+1}$  and  $\bar{\mathbf{R}}_{k+1, i}$ ,  $i = 1, \dots, k_0$ , now have more than one subblock with different parameters  $\mu$  instead of a single one. This new block structure helps us derive the optimality conditions more concisely. Clearly, we would like to achieve the optimal solution  $\mathbf{X}$  with the following structure:

$$\mathbf{X} = \begin{pmatrix} \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & & \mathbf{0} \\ \vdots & \ddots & \ddots & \mathbf{0} & \vdots \\ \vdots & & \mathbf{0} & \sigma_k \mathbf{u}_k \mathbf{v}_k^T & \mathbf{0} \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{0} \end{pmatrix},$$

where  $\|\mathbf{u}_i\|_2 = \|\mathbf{v}_i\|_2 = 1$  for  $i = 1, \dots, k$ . Padding appropriate zeros to  $\mathbf{u}_i$  and  $\mathbf{v}_i$  to construct  $\mathbf{u}_i^0 \in \mathbb{R}_+^m$  and  $\mathbf{v}_i^0 \in \mathbb{R}_+^n$  for  $i = 1, \dots, k$ , we obtain sufficient optimality conditions based on Proposition 2.7 as follows:

There exist  $\mathbf{Y}$  and  $\mathbf{Z}$  such that  $\mathbf{Y} + \mathbf{Z} = \mathbf{A}$  and

$$\mathbf{Y} = \|\mathbf{A}\|_{k,2,\theta}^* \left[ \sum_{i=1}^k \sigma_i \mathbf{u}_i^0 (\mathbf{v}_i^0)^T + \mathbf{W} \right], \quad \mathbf{Z} = \theta \|\mathbf{A}\|_{k,2,\theta}^* \mathbf{V},$$

where  $\sigma_i > 0$  for  $i = 1, \dots, k$ ,  $\sum_{i=1}^k \sigma_i^2 = 1$ ,  $\|\mathbf{W}\| \leq \min_{i=1, \dots, k} \{\sigma_i\}$ ,  $\mathbf{W} \mathbf{v}_i^0 = \mathbf{0}$ ,  $\mathbf{W}^T \mathbf{u}_i^0 = \mathbf{0}$ , for  $i = 1, \dots, k$ , and  $\|\mathbf{V}\|_\infty \leq 1$ ,  $\mathbf{V}_{ii} = \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T$ , for  $i = 1, \dots, k$ .

Since  $\mathbf{A}$  has the block structure, we can break these optimality conditions into appropriate conditions for each block. Starting with diagonal  $(i, i)$  blocks,  $i = 1, \dots, k$ , the detailed conditions are

$$(3.17) \quad \sigma_i \mathbf{u}_i \mathbf{v}_i^T + \mathbf{W}_{ii} = \lambda(\bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T + \mathbf{R}_{ii}) - \theta \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T,$$

$$(3.18) \quad \mathbf{W}_{ii}^T \mathbf{u}_i = \mathbf{0},$$

$$(3.19) \quad \mathbf{W}_{ii} \mathbf{v}_i = \mathbf{0},$$

where  $\lambda = 1/\|\mathbf{A}\|_{k,2,\theta}^*$ . For nondiagonal  $(i, j)$  blocks,  $i \neq j$  and  $i, j = 1, \dots, k$ , we obtain the following conditions:

$$(3.20) \quad \mathbf{W}_{ij} + \theta \mathbf{V}_{ij} = \lambda \mathbf{R}_{ij},$$

$$(3.21) \quad \mathbf{W}_{ij}^T \mathbf{u}_i = \mathbf{0},$$

$$(3.22) \quad \mathbf{W}_{ij} \mathbf{v}_j = \mathbf{0},$$

$$(3.23) \quad \|\mathbf{V}_{ij}\|_\infty \leq 1.$$

For  $(i, k+1)$  blocks,  $i = 1, \dots, k$ , we have that

$$(3.24) \quad \mathbf{W}_{i,k+1} + \theta \mathbf{V}_{i,k+1} = \lambda \bar{\mathbf{R}}_{i,k+1},$$

$$(3.25) \quad \mathbf{W}_{i,k+1}^T \mathbf{u}_i = \mathbf{0},$$

$$(3.26) \quad \|\mathbf{V}_{i,k+1}\|_\infty \leq 1.$$

Similarly, for  $(k+1, j)$  blocks,  $j = 1, \dots, k$ , the conditions are

$$(3.27) \quad \mathbf{W}_{k+1,j} + \theta \mathbf{V}_{k+1,j} = \lambda \bar{\mathbf{R}}_{k+1,j},$$

$$(3.28) \quad \mathbf{W}_{k+1,j} \mathbf{v}_j = \mathbf{0},$$

$$(3.29) \quad \|\mathbf{V}_{k+1,j}\|_\infty \leq 1.$$

Finally, the  $(k+1, k+1)$  block needs the following conditions:

$$(3.30) \quad \mathbf{W}_{k+1,k+1} + \theta \mathbf{V}_{k+1,k+1} = \lambda(\bar{\mathbf{B}}_{k+1,k+1} + \bar{\mathbf{R}}_{k+1,k+1}),$$

$$(3.31) \quad \|\mathbf{V}_{k+1,k+1}\|_\infty \leq 1.$$

The remaining conditions are not block separable. We still need  $\sigma_i > 0$ ,  $i = 1, \dots, k$ , and  $\sum_{i=1}^k \sigma_i^2 = 1$ . The last condition, which is  $\|\mathbf{W}\| \leq \min_{i=1,\dots,k} \{\sigma_i\}$ , can be replaced by the following sufficient conditions that are block separable by applying the fact that  $\|\mathbf{W}\|^2 \leq \sum_{i,j} \|\mathbf{W}_{ij}\|^2$ :

$$(3.32) \quad \|\mathbf{W}_{ij}\| \leq \frac{1}{k+1} \min_{i=1,\dots,k} \{\sigma_i\}, \quad i, j = 1, \dots, k+1.$$

With these sufficient block separable conditions, in order to construct  $(\mathbf{V}, \mathbf{W})$ , we now need to construct  $(\mathbf{V}_{ij}, \mathbf{W}_{ij})$  for different pairs  $(i, j)$  block by block. The block by block details are shown in the following analysis.

In the following proof, we assume that the random matrix  $\mathbf{R}$  is chosen in stages: the diagonal blocks  $\mathbf{R}_{ii}$ ,  $i = 1, \dots, k$ , are selected before the off-diagonal blocks. This allows us to treat the diagonal blocks as deterministic during the analysis of the off-diagonal blocks. This technique of staging independent random variables is by now standard in the literature; see e.g., the “golfing” analysis of the matrix completion problem by Gross [11].

**3.1. Analysis for block  $(i, i)$ ,  $i = 1, \dots, k$ .** We begin with the proof of the existence of a  $\lambda > 0$  that satisfies the optimality conditions. We then show the sufficient condition (3.32) for block  $(i, i)$ ,  $i = 1, \dots, k$ . The final condition that needs to be proved for these blocks is the positivity of  $\mathbf{u}_i$  and  $\mathbf{v}_i$ ,  $i = 1, \dots, k$ .

**3.1.1. Existence of  $\lambda^*$ .** The conditions for the  $(i, i)$  block,  $i = 1, \dots, k$ , namely, (3.17)–(3.19), indicate that  $(\sigma_i, \mathbf{u}_i, \mathbf{v}_i)$  is the dominant singular triple of the matrix  $\mathbf{L}_i = \lambda(\bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T + \mathbf{R}_{ii}) - \theta \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T$ . They also indicate that

$$(3.33) \quad \|\mathbf{W}_{ii}\| = \sigma_2(\mathbf{L}_i)$$

since (3.17)–(3.19) are equivalent to the first step of a singular value decomposition of  $\mathbf{L}_i$ .

For the rest of this analysis, it is more convenient notationally to work with  $\tau = \lambda/\theta$  rather than with  $\lambda$  directly. The condition  $\sum_{i=1}^k \sigma_i^2 = 1$  becomes

$$(3.34) \quad f(\tau) = \sum_{i=1}^k \left\| \tau(\bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T + \mathbf{R}_{ii}) - \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T \right\|^2 - \theta^{-2} = 0.$$

We will prove that there exists  $\tau^* > 0$  such that  $f(\tau^*) = 0$ . More precisely, we will focus our analysis of  $f(\tau)$  for  $\tau \in [\tau_\ell, \tau_u]$ , where  $\tau_\ell$  is given by (3.110) and  $\tau_u$  is given by (3.116) below, and prove that there exists  $\tau^* \in [\tau_\ell, \tau_u]$  such that  $f(\tau^*) = 0$ .

Letting  $\mathbf{Q}_{ij} = \mathbf{R}_{ij}/\sqrt{\phi_i \phi_j} - \mu_{ij} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T$  for  $i, j = 1, \dots, k$ , clearly,  $\mathbf{Q}_{ij}$  are  $b$ -subgaussian random matrices with independent elements. The function  $f$  can be rewritten as follows:

$$(3.35) \quad \begin{aligned} f(\tau) &= \sum_{i=1}^k \left\| \tau \bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T - (1 - \tau \phi_i \mu_{ii}) \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T + \tau \phi_i \mathbf{Q}_{ii} \right\|^2 - \theta^{-2} \\ &= \sum_{i=1}^k \left\| \mathbf{P}_i(\tau) + \tau \phi_i \mathbf{Q}_{ii} \right\|^2 - \theta^{-2}, \end{aligned}$$

where  $\mathbf{P}_i(\tau) = \tau \bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T - (1 - \tau \phi_i \mu_{ii}) \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T$ . Applying the triangle inequality, we have

$$(3.36) \quad \|\mathbf{P}_i(\tau)\| - \tau \phi_i \|\mathbf{Q}_{ii}\| \leq \|\mathbf{P}_i(\tau) + \tau \phi_i \mathbf{Q}_{ii}\| \leq \|\mathbf{P}_i(\tau)\| + \tau \phi_i \|\mathbf{Q}_{ii}\|.$$

We start the analysis with  $\|\mathbf{P}_i(\tau)\|$ . We first define the function

$$(3.37) \quad g_i(\tau; a) = \phi_i^2 \tau^2 - 2a \phi_i \tau (1 - \mu_{ii} \phi_i \tau) + (1 - \mu_{ii} \phi_i \tau)^2,$$

which is a quadratic function in  $\tau$  with any fixed parameter  $a$ . Note by (3.119) below that  $\tau_u \leq 0.3/(\phi_i \mu_{ii})$  for all  $i = 1, \dots, k$ , so  $1 - \mu_{ii} \phi_i \tau \geq 0$  and  $\tau \geq 0$  for  $\tau \in [\tau_\ell, \tau_u]$ . Therefore, provided  $a \leq 1$  and  $\tau \in [\tau_\ell, \tau_u]$ ,

$$(3.38) \quad \begin{aligned} g_i(\tau; a) &= (\phi_i \tau - (1 - \mu_{ii} \phi_i \tau))^2 + 2(1 - a) \phi_i \tau (1 - \mu_{ii} \phi_i \tau) \\ &\geq (\phi_i \tau - (1 - \mu_{ii} \phi_i \tau))^2 \\ &= (\phi_i (1 + \mu_{ii}) \tau - 1)^2 \end{aligned}$$

$$(3.39) \quad \geq 0.$$

The dominant singular triple of  $\mathbf{P}_i(\tau) = \tau \bar{\sigma}_i \bar{\mathbf{u}}_i \bar{\mathbf{v}}_i^T - (1 - \tau \phi_i \mu_{ii}) \mathbf{e}_{m_i} \mathbf{e}_{n_i}^T$  is now analyzed for a fixed  $\tau \in [\tau_\ell, \tau_u]$ . It is clear that the dominant right singular vector

lies in  $\text{span}\{\bar{\mathbf{v}}_i, \mathbf{e}_{n_i}\}$  since this is the range of  $(\mathbf{P}_i(\tau))^T$ . Letting  $\zeta_i = \|\mathbf{P}_i(\tau)\|^2$  be the square of the dominant singular value, it is clear that  $\zeta_i$  is a solution of the following eigenvector problem:

$$(\mathbf{P}_i(\tau))^T \mathbf{P}_i(\tau)(\alpha \bar{\mathbf{v}}_i + \beta \mathbf{e}_{n_i}) = \zeta_i(\alpha \bar{\mathbf{v}}_i + \beta \mathbf{e}_{n_i}).$$

Expanding and gathering multiples of  $\bar{\mathbf{v}}_i$  and  $\mathbf{e}_{n_i}$ , we obtain the  $2 \times 2$  eigenvalue problem

$$(3.40) \quad \mathbf{M}_i \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \zeta_i \begin{pmatrix} \alpha \\ \beta \end{pmatrix},$$

where  $\mathbf{M}_i$  is the  $2 \times 2$  matrix,

$$(3.41) \quad \begin{pmatrix} \tau^2 \bar{\sigma}_i^2 - \tau \bar{\sigma}_i h_i(\tau) \|\bar{\mathbf{u}}_i\|_1 \|\bar{\mathbf{v}}_i\|_1 & \tau^2 \bar{\sigma}_i^2 - \tau \bar{\sigma}_i h_i(\tau) \|\bar{\mathbf{u}}_i\|_1 n_i \\ (h_i(\tau))^2 \|\bar{\mathbf{v}}_i\|_1 m_i - \tau \bar{\sigma}_i h_i(\tau) \|\bar{\mathbf{u}}_i\|_1 & (h_i(\tau))^2 m_i n_i - \tau \bar{\sigma}_i h_i(\tau) \|\bar{\mathbf{u}}_i\|_1 \|\bar{\mathbf{v}}_i\|_1 \end{pmatrix},$$

and  $h_i(\tau) = 1 - \tau \phi_i \mu_{ii}$ ,  $i = 1, \dots, k$ . Thus,  $\zeta_i$  is a root of the equation

$$(3.42) \quad \zeta_i^2 - \text{trace}(\mathbf{M}_i) \zeta_i + \det(\mathbf{M}_i) = 0,$$

where

$$(3.43) \quad \begin{aligned} \text{trace}(\mathbf{M}_i) &= \tau^2 \bar{\sigma}_i^2 - 2\tau \bar{\sigma}_i (1 - \tau \phi_i \mu_{ii}) \|\bar{\mathbf{u}}_i\|_1 \|\bar{\mathbf{v}}_i\|_1 + (1 - \tau \phi_i \mu_{ii})^2 m_i n_i \\ &= m_i n_i [\tau^2 \phi_i^2 - 2\tau \phi_i (1 - \tau \phi_i \mu_{ii}) \delta_{u,i} \delta_{v,i} + (1 - \tau \phi_i \mu_{ii})^2] \\ &= m_i n_i g_i(\tau; \delta_{u,i} \delta_{v,i}) \end{aligned}$$

and

$$(3.44) \quad \begin{aligned} \det(\mathbf{M}_i) &= \tau^2 \bar{\sigma}_i^2 (1 - \tau \phi_i \mu_{ii})^2 (m_i - \|\bar{\mathbf{u}}_i\|_1^2) (n_i - \|\bar{\mathbf{v}}_i\|_1^2) \\ &= m_i^2 n_i^2 \tau^2 \phi_i^2 (1 - \tau \phi_i \mu_{ii})^2 (1 - \delta_{u,i}^2) (1 - \delta_{v,i}^2) \\ &\geq 0. \end{aligned}$$

Here, we have introduced notation

$$\begin{aligned} \delta_{u,i} &= \|\bar{\mathbf{u}}_i\|_1 / \sqrt{m_i}, \\ \delta_{v,i} &= \|\bar{\mathbf{v}}_i\|_1 / \sqrt{n_i} \end{aligned}$$

that we will continue to use for the remainder of the proof. It is apparent that  $\delta_{u,i} \in [\delta_u, 1]$  by (3.3) and similarly  $\delta_{v,i} \in [\delta_v, 1]$ .

Let  $\Delta$  be the discriminant of the quadratic equation (3.42), that is,

$$(3.45) \quad \Delta = \text{trace}(\mathbf{M}_i)^2 - 4 \det(\mathbf{M}_i).$$

Letting  $\Delta' = \Delta / (m_i n_i)^2$ , we have that

$$(3.46) \quad \begin{aligned} \Delta' &= \left[ \tau^2 \phi_i^2 - 2\tau \phi_i (1 - \tau \phi_i \mu_{ii}) \left( \delta_{u,i} \delta_{v,i} + \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)} \right) + (1 - \tau \phi_i \mu_{ii})^2 \right] \\ &\quad \cdot \left[ \tau^2 \phi_i^2 - 2\tau \phi_i (1 - \tau \phi_i \mu_{ii}) \left( \delta_{u,i} \delta_{v,i} - \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)} \right) + (1 - \tau \phi_i \mu_{ii})^2 \right] \\ &= g_i \left( \tau; \delta_{u,i} \delta_{v,i} + \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)} \right) \\ &\quad \cdot g_i \left( \tau; \delta_{u,i} \delta_{v,i} - \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)} \right). \end{aligned}$$



As  $1 - (\delta_{u,i}\delta_{v,i} + \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)})^2 = (\delta_{u,i}\sqrt{1 - \delta_{v,i}^2} - \delta_{v,i}\sqrt{1 - \delta_{u,i}^2})^2 \geq 0$ , the second argument to each invocation of  $g_i$  in the previous equation is less than or equal to 1. In addition, since  $\tau \in [\tau_\ell, \tau_u]$ , it follows that both evaluations of  $g_i$  yield nonnegative numbers, and therefore  $\Delta \geq 0$ .

We next claim that

$$(3.47) \quad \Delta = m_i^2 n_i^2 g_i(\tau; p_i(\tau))^2$$

for a continuous  $p_i(\tau) \in [\delta_{u,i}\delta_{v,i}, 1]$  for all  $\tau \in [\tau_\ell, \tau_u]$ . In other words, there exists a continuous  $p_i(\tau)$  in the range  $[a, 1]$  satisfying the equation

$$(3.48) \quad g_i(\tau; p_i(\tau))^2 = g_i(\tau; a + c)g_i(\tau; a - c),$$

where, for this paragraph,  $a = \delta_{u,i}\delta_{v,i}$  and  $c = \sqrt{(1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)}$ . This is proved by first treating  $p_i$  as an unknown and expanding (3.48). After simplification, the result is a quadratic equation for  $p_i$ . The facts that  $0 \leq a, c \leq 1$  and  $a + c \leq 1$  allow one to argue that the quadratic equation has a sign change over the interval  $[a, 1]$  for all  $\tau \in [0, 1/(\phi_i\mu_{ii})]$  (hence for all  $\tau \in [\tau_\ell, \tau_u]$ ). Thus, the quadratic has a unique root in this interval, which may be taken to be  $p_i$ ; it must vary continuously with the coefficients of the quadratic and hence with  $\tau$ . The details are left to the reader. In addition to  $\tau$ ,  $p_i(\tau)$  depends on  $\mu_{ii}$ ,  $\phi_i$ ,  $\delta_{u,i}$ , and  $\delta_{v,i}$ .

Thus, by the quadratic formula applied to (3.42), we can obtain  $\zeta_i$  as the larger root

$$(3.49) \quad \zeta_i = \frac{1}{2}(\text{trace}(\mathbf{M}_i(\tau)) + \sqrt{\Delta}) = m_i n_i g_i(\tau; a_i(\tau)),$$

where the second equation comes from adding (3.43) to the square root of (3.47) and noting that for any  $\tau, a, b$ ,  $(g_i(\tau; a) + g_i(\tau; b))/2 = g_i(\tau; (a + b)/2)$ . Here, we have that

$$(3.50) \quad a_i(\tau) = \frac{1}{2}(\delta_{u,i}\delta_{v,i} + p_i(\tau)).$$

By the earlier bound on  $p_i(\tau)$ , this implies  $a_i(\tau) \in [\underline{a}_i, \bar{a}_i]$ , where

$$(3.51) \quad \underline{a}_i = \delta_{u,i}\delta_{v,i}; \quad \bar{a}_i = \frac{1}{2} + \frac{1}{2}\delta_{u,i}\delta_{v,i}.$$

Clearly  $0 \leq \underline{a}_i \leq \bar{a}_i \leq 1$  for all  $i$  since  $\delta_{u,i}, \delta_{v,i} \in [0, 1]$ . Note that tighter bounds are possible by a more careful analysis of  $\Delta$ .

Since  $\zeta_i = \|\mathbf{P}_i(\tau)\|^2$ , we can then express  $\|\mathbf{P}_i(\tau)\|$  as follows:

$$(3.52) \quad \|\mathbf{P}_i(\tau)\| = \sqrt{\zeta_i} = \sqrt{m_i n_i g_i(\tau; a_i(\tau))}.$$

Next, consider again (3.38); the right-hand side is a convex quadratic function of  $\tau$  with minimizer at  $1/(\phi_i(1 + \mu_{ii}))$ . It follows from (3.112) that  $\tau_\ell \geq 2/\phi_i$  for all  $i$ . Thus, for  $\tau \in [\tau_\ell, \tau_u]$ , we have that

$$\tau \geq \frac{2}{\phi_i} \geq \frac{1}{\phi_i(1/2 + \mu_{ii})} > \frac{1}{\phi_i(1 + \mu_{ii})}.$$

Thus, the right-hand side of (3.38) is an increasing function of  $\tau$  for  $\tau \in [\tau_\ell, \tau_u]$ . We then have, for any  $\tau \in [\tau_\ell, \tau_u]$ ,

$$\begin{aligned} g_i(\tau; a) &\geq \left( \phi_i(1 + \mu_{ii}) \left( \frac{1}{\phi_i(1/2 + \mu_{ii})} \right) - 1 \right)^2 \\ &= \left( \frac{1}{1 + 2\mu_{ii}} \right)^2. \end{aligned}$$

Thus,

$$(3.53) \quad \|P_i(\tau)\| \geq \frac{\sqrt{m_i n_i}}{1 + 2\mu_{ii}}$$

for any  $\tau \in [\tau_\ell, \tau_u]$ . We also have a second lower bound that grows linearly with  $\tau$ :

$$(3.54) \quad \begin{aligned} \sqrt{g_i(\tau; a)} &\geq \phi_i(1 + \mu_{ii})\tau - 1 \\ &= \phi_i\tau/2 + (\phi_i(1/2 + \mu_{ii})\tau - 1) \end{aligned}$$

$$(3.55) \quad \geq \phi_i\tau/2,$$

where the first inequality follows from (3.38) and the other inequality is due to the fact that  $\tau\phi_i \geq 2$  as noted above. This implies

$$(3.56) \quad \begin{aligned} \|P_i(\tau)\| &\geq \sqrt{m_i n_i}\phi_i\tau/2 \\ &= \bar{\sigma}_i\tau/2. \end{aligned}$$

Next, we combine this linear lower bound on  $\|P_i(\tau)\|$  with an upper bound on  $\|Q_{ii}\|$  in order to be able to take advantage of (3.36).

CLAIM 1.  $\|Q_{ij}\| \leq (m_i n_j)^{\frac{3}{8}}$  with probability exponentially close to 1 as  $m_i, n_j \rightarrow \infty$  for all  $i, j = 1, \dots, k$ .

To establish the claim observe that  $Q_{ij}$  is random with i.i.d. elements that are  $b$ -subgaussian. Thus by Lemma 3.2(i),

$$(3.57) \quad \mathbb{P}\left(\|Q_{ij}\| \geq (m_i n_j)^{3/8}\right) \leq \exp\left(-\left(\frac{(m_i n_j)^{3/4}}{81b^2} - (\log 7)(m_i + n_j)\right)\right),$$

where  $u$  is set to be  $(m_i n_j)^{3/8}$ . The right-hand side tends to zero exponentially fast since  $(m_i n_j)^{3/4}$  asymptotically dominates  $m_i + n_j$  under the assumption that  $m_i^{1/4} \leq O(n_j^{1/2})$  and  $n_j^{1/4} \leq O(m_i^{1/2})$ , which was stated as a hypothesis in the theorem.

Then with a probability exponentially close to 1, the event in (3.57) does not happen, hence we assume  $\|Q_{ij}\| \leq (m_i n_j)^{\frac{3}{8}}$ . Focusing on the  $i = j$  case for now, this implies

$$(3.58) \quad \|Q_{ii}\| \leq \sqrt{m_i n_i}/40$$

for large  $m_i n_i$ ; since the theorem applies to the asymptotic range, we assume this inequality holds true as well. Combining the inequality (3.58) with (3.56) and (3.36), we obtain

$$(3.59) \quad \begin{aligned} \|P_i(\tau) + \tau\phi_i Q_{ii}\| &= (1 + \gamma_i(\tau)) \|P_i(\tau)\| \\ &= (1 + \gamma_i(\tau)) \sqrt{m_i n_i g_i(\tau; a_i(\tau))}. \end{aligned}$$

In the first line, we have introduced scalar  $\gamma_i(\tau)$  to stand for a quantity in the range  $[-1/20, 1/20]$  that varies continuously with  $\tau$ . This notation will be used throughout the remainder of the proof. The second line follows from (3.52). Combining (3.59) and (3.56), we conclude

$$(3.60) \quad \|P_i(\tau) + \tau\phi_i Q_{ii}\| \geq 0.47\bar{\sigma}_i\tau.$$

Finally, because  $P_i(\tau) + \tau\phi_i Q_{ii}$  is a rescaling of the right-hand side of (3.17) by  $\theta$ , we conclude that

$$(3.61) \quad \sigma_i \geq 0.47\bar{\sigma}_i\tau\theta.$$

Applying (3.59) to the formulation of  $f(\tau)$  in (3.35), we have that, for  $\tau \in [\tau_\ell, \tau_u]$ ,

$$\begin{aligned} f(\tau) &= \sum_{i=1}^k \|\mathbf{P}_i(\tau) + \tau\phi_i\mathbf{Q}_{ii}\|^2 - \theta^{-2} \\ &= \sum_{i=1}^k m_i n_i (1 + \gamma_i(\tau))^2 g_i(\tau; a_i(\tau)) - \theta^{-2} \\ &= A(\tau)\tau^2 - 2B(\tau)\tau - C(\tau). \end{aligned}$$

The third line is obtained by expanding the quadratic formula for  $g_i(\tau; a_i(\tau))$ , which results in

$$\begin{aligned} A(\tau) &= \sum_{i=1}^k (1 + \gamma_i(\tau))^2 \bar{\sigma}_i^2 (1 + 2\mu_{ii}a_i(\tau) + \mu_{ii}^2), \\ B(\tau) &= \sum_{i=1}^k (1 + \gamma_i(\tau))^2 \sqrt{m_i n_i} \bar{\sigma}_i (a_i(\tau) + \mu_{ii}), \\ C(\tau) &= \theta^{-2} - \sum_{i=1}^k (1 + \gamma_i(\tau))^2 m_i n_i. \end{aligned}$$

We will now prove that there exists  $\tau^* \in [\tau_\ell, \tau_u]$  such that  $f(\tau^*) = 0$  by applying the following lemma, which is a specific form of the intermediate theorem for “pseudo-quadratic” functions.

LEMMA 3.5. *Consider a real-valued function  $\hat{f}(\tau)$  of the form*

$$\hat{f}(\tau) = A(\tau)\tau^2 - 2B(\tau)\tau - C(\tau),$$

where  $A(\tau)$ ,  $B(\tau)$ ,  $C(\tau)$  are continuous functions of  $\tau$ . Suppose there are two triples of positive numbers  $(\underline{A}, \underline{B}, \underline{C}) < (\overline{A}, \overline{B}, \overline{C})$  (where  $<$  is understood elementwise). Define

$$(3.62) \quad \tau'_\ell = \frac{\underline{B} + \sqrt{\underline{B}^2 + \underline{A} \cdot \underline{C}}}{\underline{A}}$$

and

$$(3.63) \quad \tau'_u = \frac{\overline{B} + \sqrt{\overline{B}^2 + \overline{A} \cdot \overline{C}}}{\underline{A}}.$$

(Clearly  $\tau'_\ell < \tau'_u$ .) Suppose further that there is an interval  $[\tau_\ell, \tau_u]$  such that  $\tau_\ell \leq \tau'_\ell \leq \tau'_u \leq \tau_u$  and such that for all  $\tau \in [\tau_\ell, \tau_u]$ ,

$$(\underline{A}, \underline{B}, \underline{C}) \leq (A(\tau), B(\tau), C(\tau)) \leq (\overline{A}, \overline{B}, \overline{C}).$$

Then there exists a root  $\tau^* \in [\tau'_\ell, \tau'_u]$  (and therefore also in  $[\tau_\ell, \tau_u]$ ) such that  $\hat{f}(\tau^*) = 0$ .

*Proof.* Some simple algebra shows that  $\hat{f}(\tau'_\ell) = A(\tau'_\ell)(\tau'_\ell)^2 - 2B(\tau'_\ell)(\tau'_\ell) - C(\tau'_\ell) \leq 0$ , while  $\hat{f}(\tau'_u) = A(\tau'_u)(\tau'_u)^2 - 2B(\tau'_u)\tau'_u - C(\tau'_u) \geq 0$ , so there is a  $\tau^* \in [\tau'_\ell, \tau'_u]$  such that  $\hat{f}(\tau^*) = 0$  by the intermediate value theorem.  $\square$

In order to apply Lemma 3.5, we now define the following scalars:

$$\begin{aligned}\overline{A} &= (10/9) \sum_{i=1}^k \bar{\sigma}_i^2 (1 + 2\mu_{ii}\bar{a}_i + \mu_{ii}^2), \\ \underline{A} &= 0.90 \sum_{i=1}^k \bar{\sigma}_i^2 (1 + 2\mu_{ii}\underline{a}_i + \mu_{ii}^2), \\ \overline{B} &= (10/9) \sum_{i=1}^k \sqrt{m_i n_i} \bar{\sigma}_i (\bar{a}_i + \mu_{ii}), \\ \underline{B} &= 0.90 \sum_{i=1}^k \sqrt{m_i n_i} \bar{\sigma}_i (\underline{a}_i + \mu_{ii}), \\ \overline{C} &= (10/9) \left( \theta^{-2} - \sum_{i=1}^k m_i n_i \right), \\ \underline{C} &= (9/10) \left( \theta^{-2} - \sum_{i=1}^k m_i n_i \right).\end{aligned}$$

It is obvious that  $(0, 0) < (\underline{A}, \underline{B}) < (\overline{A}, \overline{B})$ . It follows from (3.16), (3.111), and (3.115) below that the parenthesized quantity in the definitions of  $\overline{C}, \underline{C}$  is positive,

$$\begin{aligned}\theta^{-2} - \sum_{i=1}^k m_i n_i &\geq \left( \frac{1}{4c_3(\mathbf{p})^2} - 1 \right) \sum_{i=1}^k m_i n_i \\ &= 1.2^4 (c_4(\mathbf{p}))^2 \left( \frac{\rho_m \rho_n k}{k + \rho_m \rho_n - 1} \right) \sum_{i=1}^k m_i n_i > 0,\end{aligned}$$

and hence we also have  $0 < \underline{C} < \overline{C}$ .

In addition, given the fact that  $a_i(\tau) \in [\underline{a}_i; \bar{a}_i]$  and  $\gamma_i(\tau) \in [-1/20; 1/20]$  for  $\tau \in [\tau_\ell, \tau_u]$ , this establishes for this interval that  $(\underline{A}, \underline{B}, \underline{C}) \leq (A(\tau), B(\tau), C(\tau)) \leq (\overline{A}, \overline{B}, \overline{C})$ .

We now show that  $\tau'_\ell \geq \tau_\ell$  and  $\tau'_u \leq \tau_u$ . We have that

$$\tau'_\ell = \frac{\underline{B} + \sqrt{\underline{B}^2 + \underline{A} \cdot \underline{C}}}{\underline{A}} \geq \frac{\sqrt{\underline{A} \cdot \underline{C}}}{\underline{A}}.$$

Using the facts that  $0 \leq \bar{a}_i \leq 1$ ,  $0 \leq \mu_{ii} \leq c_2(\mathbf{p}, c_0) \leq 0.08$  (see (3.118) below), and  $\theta^{-2} - \sum_{i=1}^k m_i n_i \geq 1.2^4 (c_4(\mathbf{p}))^2 \left( \frac{\rho_m \rho_n k}{k + \rho_m \rho_n - 1} \right) \sum_{i=1}^k m_i n_i > 0$  as above, we have that

$$\tau'_\ell \geq \left( \frac{\rho_m \rho_n k}{k + \rho_m \rho_n - 1} \right)^{1/2} c_4(\mathbf{p}) \left( \sum_{i=1}^k m_i n_i \right)^{1/2} \left( \sum_{i=1}^k \bar{\sigma}_i^2 \right)^{-1/2}.$$

Next, observe that

$$\left( \sum_{i=1}^k \bar{\sigma}_i^2 \right)^{-1/2} \geq k^{-1/2} \sigma_1^{-1}$$

while

$$\left( \sum_{i=1}^k m_i n_i \right)^{1/2} \geq \left( 1 + \frac{k-1}{\rho_m \rho_n} \right)^{1/2} (m_1 n_1)^{1/2}.$$

Since  $\phi_1 = \sigma_1 / \sqrt{m_1 n_1}$ , we conclude that

$$\tau'_\ell \geq c_4(\mathbf{p}) \phi_1^{-1} = \tau_\ell,$$

given the definition of  $\tau_\ell$  in (3.110).

We now consider the condition for  $\tau'_u$ . We have that

$$\tau'_u = \frac{\overline{B} + \sqrt{\overline{B}^2 + \overline{A} \cdot \overline{C}}}{\underline{A}} = \frac{\overline{B}}{\underline{A}} + \sqrt{\frac{\overline{B}^2}{\underline{A}^2} + \frac{\overline{A} \cdot \overline{C}}{\underline{A}^2}}.$$

Using the fact that  $0 \leq \overline{a}_i \leq 1$ ,  $0 \leq \mu_{ii} \leq 0.08$ , we have that

$$\begin{aligned} \frac{\overline{B}}{\underline{A}} &\leq \frac{100}{81} \cdot \left( 1.08 \sum_{i=1}^k \overline{\sigma}_i \sqrt{m_i n_i} \right) \left( \sum_{i=1}^k \overline{\sigma}_i^2 \right)^{-1} \\ &\leq \frac{4}{3} \cdot \left( \frac{1 + (k-1) \sqrt{\rho_m \rho_n}}{1 + (k-1) \rho_\sigma^{-2}} \right) \phi_1^{-1}, \end{aligned}$$

and, using also (3.16),

$$\begin{aligned} \frac{\overline{A} \cdot \overline{C}}{\underline{A}^2} &\leq \frac{16}{9} \cdot (c_3(\mathbf{p})^{-2} - 1) \left( \sum_{i=1}^k m_i n_i \right) \left( \sum_{i=1}^k \overline{\sigma}_i^2 \right)^{-1} \\ &\leq \frac{16}{9} \cdot (c_3(\mathbf{p})^{-2} - 1) \left( \frac{1 + (k-1) \rho_m \rho_n}{1 + (k-1) \rho_\sigma^{-2}} \right) \phi_1^{-2}. \end{aligned}$$

Note that  $0 < c_3(\mathbf{p}) < 1$  given its definition in (3.115). Now, combining these terms we conclude that

$$\tau'_u \leq c_5(\mathbf{p}) \phi_1^{-1} = \tau_u,$$

given the definition of  $\tau_u$  in (3.116) below with  $c_5(\mathbf{p})$  defined in (3.117). Thus applying Lemma 3.5, we prove that there exists  $\tau^* \in [\tau_\ell, \tau_u]$  such that  $f(\tau^*) = 0$ . This also means the existence of  $\lambda^* = \theta \tau^*$ . For the remainder of this proof, we will drop the asterisks and simply write these selected values as  $\tau$  and  $\lambda$ .

Since  $\|\mathbf{A}\|_{k,2,\theta}^* = 1/\lambda$ , the  $\|\cdot\|_{k,2,\theta}^*$ -norm of  $\mathbf{A}$  is already determined at this step of the proof even though the random variables  $\mathbf{R}$  for the off-diagonal blocks of  $\mathbf{A}$  are not yet chosen. (Recall that we are assuming for the purpose of this analysis that the random variables are staged and that the diagonal-block random variables are chosen before the off-diagonal blocks.) It should not be surprising that the norm can be determined even before all entries are chosen; for many norms such as the vector  $\infty$ -norm, it is possible to make small perturbations to many coordinate entries without affecting the value of the norm.

**3.1.2. Upper bound on  $\|\mathbf{W}_{ii}\|$ .** Now consider the condition (3.32) for block  $(i, i)$ ,  $i = 1, \dots, k$ . By (3.33), it suffices to show

$$(3.64) \quad \sigma_2(\mathbf{P}_i(\tau) + \tau\phi_i\mathbf{Q}_{ii}) \leq \frac{1}{k+1}\sigma_1(\mathbf{P}_j(\tau) + \tau\phi_j\mathbf{Q}_{jj})$$

for all  $j = 1, \dots, k$ . In order to analyze  $\sigma_2(\mathbf{P}_i(\tau) + \tau\phi_i\mathbf{Q}_{ii})$ , we start with  $\sigma_2(\mathbf{P}_i(\tau))$ . Since  $\mathbf{P}_i(\tau)$  has the rank of at most two,  $\bar{\zeta}_i = \sigma_2^2(\mathbf{P}_i(\tau))$  can be computed as the smaller root of the quadratic equation (3.42),  $i = 1, \dots, k$ . Using the fact that  $\zeta_i\bar{\zeta}_i = \det(\mathbf{M}_i)$ , we have that

$$(3.65) \quad \bar{\zeta}_i = \frac{m_i n_i \tau^2 \phi_i^2 (1 - \tau\phi_i \mu_{ii})^2 (1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)}{g_i(\tau; a_i(\tau))}$$

from (3.44) and (3.49).

Now we note that from standard singular value perturbation theory (see, for example, Theorem 7.4.51 from Horn and Johnson [12]) that

$$\begin{aligned} \sigma_2(\mathbf{P}_i + \tau\phi_i\mathbf{Q}_{ii}) &\leq \sigma_2(\mathbf{P}_i) + \tau\phi_i \|\mathbf{Q}_{ii}\| \\ &\leq \left( \frac{m_i n_i \tau^2 \phi_i^2 (1 - \tau\phi_i \mu_{ii})^2 (1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)}{g_i(\tau; a_i(\tau))} \right)^{1/2} + \tau\phi_i (m_i n_i)^{3/8} \\ &\equiv T_1 + T_2. \end{aligned}$$

We handle the two terms separately. Since we are interested in the asymptotic case of  $m_i, n_i \rightarrow \infty$ , we will assume

$$(3.66) \quad (m_i n_i)^{-1/8} \leq \frac{1}{10(k+1)\rho_\sigma}$$

for all  $i, j = 1, \dots, k$ .

First, we have that

$$\begin{aligned} T_1 &= \left( \frac{m_i n_i \tau^2 \phi_i^2 (1 - \tau\phi_i \mu_{ii})^2 (1 - \delta_{u,i}^2)(1 - \delta_{v,i}^2)}{g_i(\tau; a_i(\tau))} \right)^{1/2} \\ &\leq \left( \frac{m_i n_i \tau^2 \phi_i^2}{\phi_i^2 \tau^2 / 4} \right)^{1/2} \\ (3.67) \quad &= 2\sqrt{m_i n_i} \\ &\leq \frac{2\phi_j \tau \sqrt{m_j n_j}}{6(k+1)} \\ &= \frac{\tau \bar{\sigma}_j}{3(k+1)} \\ (3.68) \quad &\leq \frac{\sigma_1(\mathbf{P}_j(\tau) + \tau\phi_j\mathbf{Q}_{jj})}{3 \cdot 0.47(k+1)}. \end{aligned}$$

The inequality in the second line follows from the fact that  $0 < 1 - \tau\phi_i \mu_{ii} \leq 1$  for  $\tau \in [\tau_\ell, \tau_u]$  for the numerator and (3.55) for the denominator. The inequality in the fourth line follows from  $\tau\phi_j \geq 6(k+1)\sqrt{\rho_m \rho_n}$ , which follows from (3.112). The last

line follows from (3.60). Next, we have that

$$\begin{aligned} T_2 &= \tau\phi_i(m_i n_i)^{3/8} \\ &= \tau\bar{\sigma}_i/(m_i n_i)^{1/8} \\ &\leq \tau\bar{\sigma}_j/(10(k+1)) \\ &\leq \frac{\sigma_1(\mathbf{P}_j(\tau) + \tau\phi_j \mathbf{Q}_{jj})}{10 \cdot 0.47(k+1)}, \end{aligned}$$

where the third line follows from (3.66) and the fourth again from (3.60). This inequality and (3.68) together establish (3.64).

**3.1.3. Positivity of  $\mathbf{u}_i$  and  $\mathbf{v}_i$ .** The final condition for the  $(i, i)$  block is the positivity of singular vectors. We will show that with high probability, the matrix  $\mathbf{S}_i = (\mathbf{P}_i(\tau) + \tau\phi_i \mathbf{Q}_{ii})^T (\mathbf{P}_i(\tau) + \tau\phi_i \mathbf{Q}_{ii})$  is positive, which implies the positivity of the right singular vector. (At the end of this subsection we consider the left singular vector.) Clearly,  $\mathbf{S}_i = \mathbf{S}_i^1 + \mathbf{S}_i^2 + \mathbf{S}_i^3 + \mathbf{S}_i^4$ , where  $\mathbf{S}_i^1 = (\mathbf{P}_i(\tau))^T \mathbf{P}_i(\tau)$ ,  $\mathbf{S}_i^2 = \tau\phi_i (\mathbf{P}_i(\tau))^T \mathbf{Q}_{ii}$ ,  $\mathbf{S}_i^3 = \tau\phi_i \mathbf{Q}_{ii}^T \mathbf{P}_i(\tau)$ , and  $\mathbf{S}_i^4 = \tau^2 \phi_i^2 \mathbf{Q}_{ii}^T \mathbf{Q}_{ii}$ . Start with  $\mathbf{S}_i^1$ . Recall  $\delta_{u,i} = \mathbf{e}_{m_i}^T \mathbf{u}_i / \sqrt{m_i}$  and  $\bar{\sigma}_i = \phi_i \sqrt{m_i n_i}$ . Letting  $\bar{S}_i^1(l, j) = S_i^1(l, j) / [(1 - \tau\phi_i \mu_{ii})^2 m_i]$ , we then have

$$\begin{aligned} \bar{S}_i^1(l, j) &= \sqrt{n_i} \max\{\bar{v}_{i,l}, \bar{v}_{i,j}\} \psi_i \left( \sqrt{n_i} \min\{\bar{v}_{i,l}, \bar{v}_{i,j}\} \psi_i - \frac{\delta_{u,i}(\bar{v}_{i,l} + \bar{v}_{i,j})}{\max\{\bar{v}_{i,l}, \bar{v}_{i,j}\}} \right) + 1 \\ (3.69) \quad &\geq \sqrt{n_i} \max\{\bar{v}_{i,l}, \bar{v}_{i,j}\} \psi_i (\xi_v \psi_i - 2) + 1, \end{aligned}$$

where we let  $\psi_i$  denote  $\tau\phi_i / (1 - \tau\phi_i \mu_{ii})$  for the remainder of the analysis of the positivity condition, and where  $\xi_v$  was defined by (3.6).

From (3.113),  $\tau_\ell \geq 4/(\xi_v \phi_i)$ . Since  $\tau \in [\tau_\ell, \tau_u]$ , it is clear that  $\tau\phi_i \geq 4/\xi_v$  (and similarly,  $\tau\phi_i \geq 4/\xi_u$ ) for all  $i = 1, \dots, k$ . Thus, since  $0 < 1 - \mu_{ii}\tau\phi_i \leq 1$ , we also conclude  $\psi_i \geq 4/\xi_v$  and hence  $\psi_i \xi_v - 2 \geq \psi_i \xi_v / 2$ . Substituting into (3.69) yields

$$\begin{aligned} S_i^1(l, j) &\geq (1 - \tau\phi_i \mu_{ii})^2 m_i [\sqrt{n_i} \max\{\bar{v}_{i,l}, \bar{v}_{i,j}\} \psi_i^2 \xi_v / 2 + 1] \\ (3.70) \quad &\geq (1 - \tau\phi_i \mu_{ii})^2 m_i (\psi_i^2 \xi_v^2 / 2 + 1) \end{aligned}$$

for  $l, j = 1, \dots, n_i$ .

Now considering the matrix  $\mathbf{S}_i^2$ , we have that

$$\begin{aligned} S_i^2(l, j) &= \tau\phi_i \sum_{s=1}^{m_i} (\tau\bar{\sigma}_i \bar{v}_{i,l} \bar{u}_{i,s} - (1 - \tau\phi_i \mu_{ii})) Q_{ii}(s, j) \\ (3.71) \quad &= \tau^2 \phi_i T_1 - \tau\phi_i T_2, \end{aligned}$$

where

$$\begin{aligned} T_1 &\equiv \sum_{s=1}^{m_i} (\bar{\sigma}_i \bar{v}_{i,l} \bar{u}_{i,s} + \phi_i \mu_{ii}) Q_{ii}(s, j), \\ T_2 &\equiv \sum_{s=1}^{m_i} Q_{ii}(s, j). \end{aligned}$$

According to Lemma 3.1,  $T_1$  and  $T_2$  are both subgaussian random variables with parameters  $b_1 \equiv b \|\bar{\sigma}_i \bar{v}_{i,l} \mathbf{u}_i + \phi_i \mu_{ii} \mathbf{e}_{m_i}\|$  and  $b_2 \equiv b \sqrt{m_i}$ , respectively. We can derive



an upper bound on  $b_1$ :

$$\begin{aligned}
 b_1 &= b \|\bar{\sigma}_i \bar{v}_{i,l} \bar{\mathbf{u}}_i + \phi_i \mu_{ii} \mathbf{e}_{m_i}\| \\
 &\leq b \bar{\sigma}_i \bar{v}_{i,l} \|\bar{\mathbf{u}}_i\| + b \phi_i \mu_{ii} \|\mathbf{e}_{m_i}\| \\
 &= b \phi_i (m_i n_i)^{1/2} \bar{v}_{i,l} + b \phi_i \mu_{ii} \sqrt{m_i} \\
 (3.72) \quad &\leq b \phi_i \sqrt{m_i} (\pi_v + \mu_{ii}),
 \end{aligned}$$

where the third line uses the definition  $\phi_i = \sigma_i / (m_i n_i)^{1/2}$  and  $\|\bar{\mathbf{u}}_i\| = 1$ , while the fourth line uses (3.8).

Considering the  $T_1$  term first, let us determine the probability that the negative of  $\tau^2 \phi_i T_1$  exceeds  $1/6$  times the lower bound given by (3.70):

$$\begin{aligned}
 \mathbb{P} \left( \tau^2 \phi_i T_1 \leq -\frac{(1 - \tau \phi_i \mu_{ii})^2 m_i \psi_i^2 \xi_v^2}{12} \right) &= \mathbb{P} \left( T_1 \leq -\frac{m_i \phi_i \xi_v^2}{12} \right) \\
 (3.73) \quad &\leq \exp \left( -\frac{m_i \xi_v^4}{288 b^2 (\pi_v + \mu_{ii})^2} \right),
 \end{aligned}$$

where the first line is obtained by dividing both sides by  $\tau^2 \phi_i$  and substituting the definition of  $\psi_i$ , while the second line is from (3.2) with  $t = m_i \phi_i \xi_v^2 / 12$  and the “ $b$ ” of (3.2) given by (3.72).

Now let us consider the probability that the negative of  $\tau \phi_i T^2$  exceeds the same quantity:

$$\begin{aligned}
 \mathbb{P} \left( \tau \phi_i T_2 \leq -\frac{(1 - \tau \phi_i \mu_{ii})^2 m_i \psi_i^2 \xi_v^2}{12} \right) &= \mathbb{P} \left( T_2 \leq -\frac{m_i \tau \phi_i \xi_v^2}{12} \right) \\
 &\leq \mathbb{P} \left( T_2 \leq -\frac{m_i \xi_v}{3} \right) \\
 (3.74) \quad &\leq \exp \left( -\frac{m_i \xi_v^2}{18 b^2} \right),
 \end{aligned}$$

where for the first line we again use  $\psi_i = \tau \phi_i / (1 - \tau \phi_i \mu_{ii})$ , for the second  $\tau \phi_i \geq 4/\xi_v$  derived above. The third uses (3.2) with  $t = m_i \xi_v / 3$  and the subgaussian parameter given by  $b_2$  above.

Combining (3.71), (3.73), and (3.74) via the union bound yields

$$\mathbb{P} \left( S_i^2(l, j) \leq -\frac{(1 - \tau \phi_i \mu_{ii})^2 m_i \psi_i^2 \xi_v^2}{6} \right) \leq \exp \left( -\frac{m_i \xi_v^4}{288 b^2 (\pi_v + \mu_{ii})^2} \right) + \exp \left( -\frac{m_i \xi_v^2}{18 b^2} \right). \quad (3.75)$$

Note that  $\mathbf{S}_i^3 = (\mathbf{S}_i^2)^T$ , which means the analysis is the same.

For the matrix  $\mathbf{S}_i^4$ , it is clear that  $S_i^4(l, j) = \tau^2 \phi_i^2 [(\mathbf{Q}_{ii}(:, l))^T \mathbf{Q}_{ii}(:, j)]$ , where the square-bracketed factor is the inner product of two independent  $b$ -subgaussian random vector for all  $l \neq j$ . (Note that when  $l = j$ ,  $S_i^4(l, j) \geq 0$  so there is nothing to analyze.) We again bound the probability that the negative of this term exceeds  $1/3$  times the lower bound given by (3.70):

$$\begin{aligned}
 \mathbb{P} \left( S_i^4(l, j) \leq -\frac{(1 - \tau \phi_i \mu_{ii})^2 m_i \psi_i^2 \xi_v^2}{6} \right) &= \mathbb{P} \left( (\mathbf{Q}_{ii}(:, l))^T \mathbf{Q}_{ii}(:, j) \leq -\frac{m_i \xi_v^2}{6} \right) \\
 (3.76) \quad &\leq \exp \left( -m_i \cdot \min \left( \frac{\xi_v^4}{576 e^2 b^4}, \frac{\xi_v^2}{24 e b^2} \right) \right),
 \end{aligned}$$

where, for the second line, we applied Lemma 3.3 with  $t = m_i \xi_v^2/6$  and  $n = m_i$ . Combining (3.70), (3.75), and (3.76), we have that

$$(3.77) \quad \mathbb{P} \left( \min_{l,j} S(l,j) \leq 0 \right) \leq n_i(n_i - 1) \cdot \left[ \exp \left( \frac{-m_i \xi_v^4}{288b^2(\pi_v^2 + 1)} \right) + (1/2) \exp \left( -m_i \cdot \min \left( \frac{\xi_v^4}{576e^2b^4}, \frac{\xi_v^2}{24eb^2} \right) \right) \right].$$

For the left singular vector, define the matrix,

$$\mathbf{T}_i = (\mathbf{P}_i(\tau) + \tau\phi_i \mathbf{Q}_{ii})(\mathbf{P}_i(\tau) + \tau\phi_i \mathbf{Q}_{ii})^T.$$

The analogous analysis (i.e., writing  $\mathbf{T}_i = \mathbf{T}_i^1 + \mathbf{T}_i^2 + \mathbf{T}_i^3 + \mathbf{T}_i^4$  as above and analyzing the four terms separately) yields

$$(3.78) \quad \mathbb{P} \left( \min_{l,j} T(l,j) \leq 0 \right) \leq m_i(m_i - 1) \cdot \left[ \exp \left( \frac{-n_i \xi_u^4}{288b^2(\pi_u^2 + 1)} \right) + (1/2) \exp \left( -n_i \cdot \min \left( \frac{\xi_u^4}{576e^2b^4}, \frac{\xi_u^2}{24eb^2} \right) \right) \right].$$

**3.2. Analysis for block  $(i, j)$ ,  $i \neq j$ ,  $i, j = 1, \dots, k$ .** We now consider the off-diagonal  $(i, j)$  block,  $i \neq j$ ,  $i, j = 1, \dots, k$ . Recall our notation:  $\mathbf{u}_i, \mathbf{v}_i$  stand for the unit-norm dominant left and right singular vectors, respectively, of the right-hand side of (3.17) or, equivalently, of  $\mathbf{P}_i(\tau) + \tau\phi_i \mathbf{Q}_{ii}$ .

Let us consider the following construction:

$$\mathbf{V}_{ij} = \tau \left( \frac{\mathbf{e}_{m_i}(\mathbf{u}_i^T \mathbf{R}_{ij})}{\|\mathbf{u}_i\|_1} + \frac{(\mathbf{R}_{ij} \mathbf{v}_j) \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} - \frac{\mathbf{u}_i^T \mathbf{R}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T \right).$$

The matrix  $\mathbf{W}_{ij} = \lambda \mathbf{R}_{ij} - \theta \mathbf{V}_{ij}$  clearly satisfies two orthogonal requirements, (3.21) and (3.22). We need to find the conditions so that  $\|\mathbf{W}_{ij}\| \leq \frac{1}{k+1} \min_{i=1, \dots, k} \sigma_i$  and  $\|\mathbf{V}_{ij}\|_\infty \leq 1$ .

**3.2.1. Upper bound on  $\|\mathbf{V}_{ij}\|_\infty$ .** We have that

$$|V_{ij}(s, t)| \leq \tau \left( \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij}(:, t)}{\|\mathbf{u}_i\|_1} \right| + \left| \frac{\mathbf{R}_{ij}(s, :) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| + \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| \right).$$

In order to show  $\|\mathbf{V}_{ij}\|_\infty \leq 1$  with high probability, we will show the sufficient condition that all probabilities,

$$(3.79) \quad \mathbb{P} \left( \tau \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij}(:, t)}{\|\mathbf{u}_i\|_1} \right| > \frac{1}{3} \right),$$

$$(3.80) \quad \mathbb{P} \left( \tau \left| \frac{\mathbf{R}_{ij}(s, :) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > \frac{1}{3} \right),$$

$$(3.81) \quad \mathbb{P} \left( \tau \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| > \frac{1}{3} \right),$$

are exponentially small.

Since  $\tau \in [\tau_\ell, \tau_u]$ , it is clear that  $\tau \sqrt{\phi_i \phi_j} \leq 0.3/\mu_{ij}$  by (3.120). Thus we have that

$$\mathbb{P} \left( \tau \left| \frac{\mathbf{R}_{ij}(s, :)\mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > \frac{1}{3} \right) \leq \mathbb{P} \left( \tau \left| \frac{(\mathbf{R}_{ij}(s, :) - \mu_{ij} \sqrt{\phi_i \phi_j} \mathbf{e}_{n_j}^T) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > \frac{1}{30} \right).$$

Thus, to analyze (3.80), it suffices to show that the probability on the right-hand side of the preceding inequality is exponentially small. Since  $\|\mathbf{v}_j\| = 1$ ,  $((\phi_i \phi_j)^{-1/2} \mathbf{R}_{ij}(s, :) - \mu_{ij} \mathbf{e}_{n_j}^T) \mathbf{v}_j$  is a  $b$ -subgaussian random variable by Lemma 3.1. (Note that  $\mathbf{v}_j$  depends on the  $(j, j)$  diagonal block of  $\mathbf{A}$ , which in turn depends on  $\mathbf{R}_{jj}$  and hence is random. However, recall also that we have assumed that the random variables in the block diagonals of  $\mathbf{R}$  are chosen before the off-diagonal blocks, so that  $\mathbf{v}_j$  may be considered as a deterministic quantity when analyzing  $\mathbf{R}_{ij}$ .)

By (3.2), we have that

$$(3.82) \quad \mathbb{P} \left( \tau \left| \frac{(\mathbf{R}_{ij}(s, :) - \mu_{ij} \sqrt{\phi_i \phi_j} \mathbf{e}_{n_j}^T) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > \frac{1}{30} \right) \leq 2 \exp \left( -\frac{0.1^2 \|\mathbf{v}_j\|_1^2}{18b^2 \tau^2 \phi_i \phi_j} \right).$$

We now must show that the probability on the right-hand side of (3.82) is exponentially small. First, we observe that

$$\begin{aligned} \tau^2 \phi_i \phi_j &\leq (\tau'_u)^2 \phi_i \phi_j \\ &\leq c_5(\mathbf{p})^2 \phi_i \phi_j / \phi_1^2 \\ &\leq c_6(\mathbf{p}), \end{aligned}$$

where  $c_6(\mathbf{p}) = c_5(\mathbf{p})^2 \rho_\sigma^2 \rho_m \rho_n$  with  $c_5(\mathbf{p})$  defined in (3.117) below. This follows from the fact that, for any  $i, j = 1, \dots, k$ ,

$$(3.83) \quad \phi_i / \phi_j = (\bar{\sigma}_i / \bar{\sigma}_j) \sqrt{m_j / m_i} \sqrt{n_j / n_i} \leq \rho_\sigma \rho_m^{1/2} \rho_n^{1/2}.$$

We now provide a lower bound on  $\|\mathbf{v}_j\|_1$ . We start with the right singular vector  $\hat{\mathbf{v}}(\mathbf{P}_j(\tau))$  of the matrix  $\mathbf{P}_j(\tau)$ . As noted prior to (3.40), this singular vector may be written as  $\hat{\alpha}_j \bar{\mathbf{v}}_j + \hat{\beta}_j \mathbf{e}_{n_j}$ . Let  $\mathbf{v}(\mathbf{P}_j(\tau))$  be the rescaling of  $\hat{\mathbf{v}}(\mathbf{P}_j(\tau))$  with the scale chosen so that  $\mathbf{v}(\mathbf{P}_j(\tau)) = \alpha_j \bar{\mathbf{v}}_j + \mathbf{e}_{n_j}$  (i.e.,  $\beta_j = 1$ ). Then we can obtain the value of  $\alpha_j$  using the second equation obtained from (3.40) (see also Lemma 4.5 in [6]), and simplifying by substituting (3.37) yields

$$(3.84) \quad \alpha_j = \frac{\sqrt{n_j}}{h_i(\tau)} \cdot \frac{\tau \phi_j [\tau \phi_j - (2a_j(\tau) - \delta_{u,i} \delta_{v,i}) h_j(\tau)]}{\delta_{v,j} h_j(\tau) - \tau \phi_j \delta_{u,j}},$$

where  $h_j(\tau) = 1 - \tau \phi_j \mu_{jj}$ , which lies in  $[0.7, 1]$  since  $\tau \leq \tau_u$ , and  $a_j(\tau)$  is defined as in (3.50). (Note that the scaling  $\beta_j = 1$  is valid only if the denominator of the above fraction is nonzero, which we shall show next.) Observe that the square-bracketed quantity in the second numerator is nonnegative and at least  $\tau \phi_j - 2$  since  $a_j \leq 1$  and  $\tau \leq \tau_u$ .

Using the facts that  $\delta_u \leq \delta_{u,i} \leq 1$  and  $\delta_v \leq \delta_{v,i} \leq 1$  we conclude from (3.114) that  $\tau \phi_j \geq 2 + 2\delta_{u,j}/\delta_{v,j}$  and  $\tau \phi_j \geq 2 + 2\delta_{v,j}/\delta_{u,j}$  for all  $j = 1, \dots, k$  whenever  $\tau \geq \tau_\ell$ .

Now, ignoring the additive term of 2 for a moment, this assumption implies that the second denominator is negative and no more than  $\tau \phi_j \delta_{u,j}$  in absolute value. Thus we have that

$$\alpha_j \leq -\frac{\sqrt{n_j}(\tau \phi_j - 2)}{\delta_{u,j}}.$$

As noted in the previous paragraph  $\tau\phi_j - 2 \geq 2\delta_{u,j}/\delta_{v,j}$ , hence

$$(3.85) \quad \alpha_j \leq -2\sqrt{n_j}/\delta_{v,j}.$$

Now we write the 1- and 2-norms of  $\mathbf{v}(P_j)$  in terms of  $\alpha_j$  and the other parameters. Starting with the 1-norm,

$$\begin{aligned} \|\mathbf{v}(P_j)\|_1 &= \|\alpha_j \bar{\mathbf{v}}_j + \mathbf{e}_{n_j}\|_1 \\ &\geq \|\alpha_j \bar{\mathbf{v}}_j\|_1 - n_j \\ &= |\alpha_j| \sqrt{n_j} \delta_{v,j} - n_j \\ &\geq |\alpha_j| \sqrt{n_j} \delta_{v,j} / 2, \end{aligned}$$

where, to obtain the last line, we used the fact that  $|\alpha_j| \sqrt{n_j} \delta_{v,j} / 2 \geq n_j$ , a consequence of (3.85). Also,  $\|\mathbf{v}(P_j)\| \leq |\alpha_j| + \sqrt{n_j}$  by the triangle inequality. Thus, we conclude that

$$\begin{aligned} \|\hat{\mathbf{v}}(P_j)\|_1 &= \frac{\|\mathbf{v}(P_j)\|_1}{\|\mathbf{v}(P_j)\|} \\ &\geq \frac{|\alpha_j| \sqrt{n_j} \delta_{v,j} / 2}{|\alpha_j| + \sqrt{n_j}} \\ &= \frac{\sqrt{n_j} \delta_{v,j}}{2(1 + \sqrt{n_j}/|\alpha_j|)} \\ &\geq \frac{\sqrt{n_j} \delta_{v,j}}{2(1 + \delta_{v,j}/2)} \\ &\geq \frac{\sqrt{n_j} \delta_{v,j}}{3}. \end{aligned} \quad (3.86)$$

Next, we observe by the triangle inequality that

$$\begin{aligned} \|\mathbf{v}_j\|_1 &\geq \|\hat{\mathbf{v}}(\mathbf{P}_j)\|_1 - \|\hat{\mathbf{v}}(\mathbf{P}_j) - \mathbf{v}_j\|_1 \\ (3.87) \quad &\geq \|\hat{\mathbf{v}}(\mathbf{P}_j)\|_1 - \sqrt{n_j} \|\hat{\mathbf{v}}(\mathbf{P}_j) - \mathbf{v}_j\|. \end{aligned}$$

We will use Wedin's theorem on perturbation of singular vectors (see Doan and Vavasis [6] and references therein for details) to analyze the final norm in the above inequality since  $\mathbf{v}_j$  is the leading singular vector of  $\mathbf{P}_j(\tau) + \tau\phi_j \mathbf{Q}_{jj}$ , while  $\hat{\mathbf{v}}(\mathbf{P}_j)$  is the leading singular vector of  $\mathbf{P}_j(\tau)$ .

For Wedin's theorem, we choose  $\mathbf{A} = \mathbf{P}_j(\tau)$ ,  $\mathbf{T} = \tau\phi_j \mathbf{Q}_{jj}$ , and  $\mathbf{B} = \mathbf{A} + \mathbf{T}$ . We have  $\|\mathbf{T}\| \leq \tau\phi_j(m_j n_j)^{3/8}$ . In addition,

$$\begin{aligned} \sigma_1(\mathbf{B}) &\geq \sigma_1(\mathbf{A}) - \sigma_1(\mathbf{T}) \\ &\geq \sqrt{m_j n_j g_j(\tau; a_j)} - \tau\phi_j(m_j n_j)^{3/8}, \end{aligned}$$

where the second line is obtained from (3.52). Finally, using (3.65),

$$\sigma_2(\mathbf{A}) = \tau\phi_j h_j(\tau) \left( \frac{m_j n_j (1 - \delta_{u,j}^2)(1 - \delta_{v,j}^2)}{g_j(\tau; a_j)} \right)^{1/2}.$$

Therefore,  $\sin \theta(\mathbf{v}_j, \hat{\mathbf{v}}(\mathbf{P}_j(\tau)))$  is upper bounded by

$$\frac{\tau\phi_j(m_j n_j)^{3/8}}{\sqrt{m_j n_j g_j(\tau; a_j)} - \tau\phi_j(m_j n_j)^{3/8} - \tau\phi_j h_j(\tau) \left( \frac{m_j n_j (1 - \delta_{u,j}^2)(1 - \delta_{v,j}^2)}{g_j(\tau; a_j)} \right)^{1/2}},$$

or equivalently,

$$\frac{(m_j n_j)^{-1/8}}{\sqrt{g_j(\tau; a_j)/(\tau \phi_j) - (m_j n_j)^{-1/8} - h_j(\tau) \left( \frac{(1-\delta_{u,j}^2)(1-\delta_{v,j}^2)}{g_j(\tau; a_j)} \right)^{1/2}}}.$$

Observe that the numerator tends to zero like  $(m_j n_j)^{-1/8}$ , while the denominator does not depend on  $m_j n_j$  (except for a vanishing term). Furthermore, the denominator is positive; this follows from the fact that the first term in the denominator is at least 0.5 by (3.55), whereas the last term is at most  $1/\sqrt{5}$  again by (3.55) and the fact that  $\phi_i \tau \geq 5$  thanks to (3.112).

This shows that

$$(3.88) \quad \|\mathbf{v}_j - \hat{\mathbf{v}}(\mathbf{P}_j(\tau))\|_2 \leq O\left((m_j n_j)^{-\frac{1}{8}}\right).$$

Combining (3.86), (3.87), and (3.88), we can then pick a constant less than  $1/3$ , say, 0.3, and claim that

$$(3.89) \quad \|\mathbf{v}_j\|_1 \geq 0.3\delta_{v,j}\sqrt{n_j} \geq 0.3\delta_v\sqrt{n_j},$$

as long as  $m_j, n_j$  are large. Combining this bound with (3.82), we can claim that the probability (3.80) is exponential small:

$$(3.90) \quad \mathbb{P}\left(\tau \left| \frac{\mathbf{R}_{ij}(s, \cdot) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > \frac{1}{3}\right) \leq 2 \exp\left(-\frac{0.1^2 0.3^2 \delta_v^2 n_j}{18b^2 c_6(\mathbf{p})}\right).$$

Similarly, the first probability (3.79) can also be proved to be exponentially small using the analogous lower bound of  $\|\mathbf{u}_i\|_1$ :

$$(3.91) \quad \|\mathbf{u}_i\|_1 \geq 0.3\delta_u\sqrt{m_i}.$$

The bound for the first probability can therefore be written as follows:

$$(3.92) \quad \mathbb{P}\left(\tau \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij}(\cdot, t)}{\|\mathbf{u}_i\|_1} \right| > \frac{1}{3}\right) \leq 2 \exp\left(-\frac{0.1^2 0.3^2 \delta_u m_i}{18b^2 c_6(\mathbf{p})}\right).$$

For the third probability (3.81), we again use the fact that  $\tau\sqrt{\phi_i \phi_j} \leq 0.3/\mu_{ij}$  since  $\tau \leq \tau_u$ :

$$\mathbb{P}\left(\tau \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| > \frac{1}{3}\right) \leq \mathbb{P}\left(\tau \left| \frac{\mathbf{u}_i^T (\mathbf{R}_{ij} - \mu_{ij} \sqrt{\phi_i \phi_j} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T) \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| > \frac{1}{30}\right),$$

where  $\mathbf{u}_i^T (\mathbf{R}_{ij} / \sqrt{\phi_i \phi_j} - \mu_{ij} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T) \mathbf{v}_j$  is a  $b$ -subgaussian random variable due to the fact that  $\|\mathbf{u}_i\|_2 = \|\mathbf{v}_j\|_2 = 1$ . We again can bound this probability using the lower bounds of  $\|\mathbf{u}_i\|_1$  and  $\|\mathbf{v}_j\|_1$  as follows:

$$(3.93) \quad \mathbb{P}\left(\tau \left| \frac{\mathbf{u}_i^T \mathbf{R}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| > \frac{1}{3}\right) \leq 2 \exp\left(-\frac{0.1^2 0.3^4 \delta_u^2 \delta_v^2 m_i n_j}{18b^2 c_6(\mathbf{p})}\right).$$

Combining (3.90), (3.92), (3.93), we obtain the following tail bound:

$$(3.94) \quad \begin{aligned} \mathbb{P}(\|\mathbf{V}_{ij}\|_\infty > 1) &\leq 2 \exp\left(-\frac{0.1^2 0.3^2 \delta_v^2 n_j}{18b^2 c_6(\mathbf{p})}\right) \\ &\quad + 2 \exp\left(-\frac{0.1^2 0.3^2 \delta_u^2 m_j}{18b^2 c_6(\mathbf{p})}\right) \\ &\quad + 2 \exp\left(-\frac{0.1^2 0.3^4 \delta_u^2 \delta_v^2 m_i n_j}{18b^2 c_6(\mathbf{p})}\right). \end{aligned}$$

**3.2.2. Upper bound on  $\|\mathbf{W}_{ij}\|$ .** The second constraint for this type of block is  $\|\mathbf{W}_{ij}\| \leq \frac{1}{k+1} \min_{i=1,\dots,k} \sigma_i$ . Using the fact that  $\mathbf{Q}_{ij} = \mathbf{R}_{ij}/\sqrt{\phi_i\phi_j} - \mu_{ij}\mathbf{e}_{m_i}\mathbf{e}_{n_j}^T$ , we have that

$$\mathbf{W}_{ij} = \tau\theta\sqrt{\phi_i\phi_j} \left( \mathbf{Q}_{ij} - \frac{\mathbf{e}_{m_i}\mathbf{u}_i^T \mathbf{Q}_{ij}}{\|\mathbf{u}_i\|_1} - \frac{\mathbf{Q}_{ij}\mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} + \frac{\mathbf{u}_i^T \mathbf{Q}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \mathbf{e}_{m_i} \mathbf{e}_{n_j}^T \right).$$

We will establish that  $\|\mathbf{W}_{ij}\| \leq \frac{1}{k+1} \min_{i=1,\dots,k} \sigma_i$  by showing that

$$(3.95) \quad \tau\theta\sqrt{\phi_i\phi_j} \left\| \frac{\mathbf{Q}_{ij}}{2} - \frac{\mathbf{e}_{m_i}\mathbf{u}_i^T \mathbf{Q}_{ij}}{\|\mathbf{u}_i\|_1} \right\| \leq \frac{1}{3(k+1)} \min_{i=1,\dots,k} \sigma_i,$$

$$(3.96) \quad \tau\theta\sqrt{\phi_i\phi_j} \left\| \frac{\mathbf{Q}_{ij}}{2} - \frac{\mathbf{Q}_{ij}\mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} \right\| \leq \frac{1}{3(k+1)} \min_{i=1,\dots,k} \sigma_i,$$

$$(3.97) \quad \tau\theta\sqrt{\phi_i\phi_j m_i n_j} \left| \frac{\mathbf{u}_i^T \mathbf{Q}_{ij} \mathbf{v}_j}{\|\mathbf{u}_i\|_1 \|\mathbf{v}_j\|_1} \right| \leq \frac{1}{3(k+1)} \min_{i=1,\dots,k} \sigma_i.$$

Given that  $m_i, n_i \rightarrow \infty$  for all  $i = 1, \dots, k$ , we make the assumption

$$(3.98) \quad (m_j n_i)^{-1/8} \leq \frac{.47}{3(k+1)(\bar{\rho}_m \bar{\rho}_n)^{1/8} \rho_\sigma c_7(\mathbf{p})}$$

for all  $i, j = 1, \dots, k$ , where we introduce

$$(3.99) \quad c_7(\mathbf{p}) = \max \left\{ \frac{1}{2} + \frac{1}{0.3\delta_u}, \frac{1}{2} + \frac{1}{0.3\delta_v}, \frac{1}{0.3^2\delta_u\delta_v} \right\}.$$

Now, inequality (3.95) is derived as follows:

$$\begin{aligned} \left\| \frac{\mathbf{Q}_{ij}}{2} - \frac{\mathbf{e}_{m_i}\mathbf{u}_i^T \mathbf{Q}_{ij}}{\|\mathbf{u}_i\|_1} \right\| &\leq \|\mathbf{Q}_{ij}\| \cdot \left\| \frac{\mathbf{I}}{2} - \frac{\mathbf{e}_{m_i}\mathbf{u}_i^T}{\|\mathbf{u}_i\|_1} \right\| \\ &\leq (m_i n_j)^{3/8} \cdot \left( \frac{1}{2} + \frac{\sqrt{m_i}}{\|\mathbf{u}_i\|_1} \right) \\ &\leq (m_i n_j)^{3/8} \cdot \left( \frac{1}{2} + \frac{1}{0.3\delta_{u,i}} \right). \end{aligned}$$

The first line uses submultiplicativity of the 2-norm since we have that

$$\mathbf{Q}_{ij}/2 - \mathbf{e}_{m_i}\mathbf{u}_i^T \mathbf{Q}_{ij}/\|\mathbf{u}_i\|_1 = (\mathbf{I}/2 - \mathbf{e}_{m_i}\mathbf{u}_i^T/\|\mathbf{u}_i\|_1)\mathbf{Q}_{ij}.$$

The second uses the triangle inequality, and the third uses (3.91). Multiply by the scalar  $\tau\theta\sqrt{\phi_i\phi_j}$  and let  $l = 1, \dots, k$  be arbitrary:

$$\begin{aligned} \tau\theta\sqrt{\phi_i\phi_j} \left\| \frac{\mathbf{Q}_{ij}}{2} - \frac{\mathbf{e}_{m_i}\mathbf{u}_i^T \mathbf{Q}_{ij}}{\|\mathbf{u}_i\|_1} \right\| &\leq \tau\theta\sqrt{\phi_i\phi_j} (m_i n_j)^{3/8} \cdot \left( \frac{1}{2} + \frac{1}{0.3\delta_{u,i}} \right) \\ &= \tau\theta\sqrt{\bar{\sigma}_i \bar{\sigma}_j} \frac{m_i^{1/8} n_j^{1/8}}{m_j^{1/4} n_i^{1/4}} \cdot \left( \frac{1}{2} + \frac{1}{0.3\delta_{u,i}} \right) \\ &\leq \frac{.47\tau\theta}{3(k+1)} \bar{\sigma}_l \\ &\leq \frac{\sigma_l}{3(k+1)}. \end{aligned}$$

The third line follows from (3.98) and the last from (3.61). Inequality (3.96) is established using the same argument. Finally, (3.97) is established by a similar argument starting from the inequality  $|\mathbf{u}_i^T \mathbf{Q}_{ij} \mathbf{v}_j| \leq \|\mathbf{Q}_{ij}\| \leq (m_i n_j)^{3/8}$ .

**3.3. Analysis for block  $(k+1, j)$ ,  $j = 1, \dots, k$ .** We now consider the  $(k+1, j)$  block. Similar to the above approach, we will construct the following matrix  $\mathbf{V}_{k+1, j}$ :

$$\mathbf{V}_{k+1, j} = \tau \frac{\bar{\mathbf{R}}_{k+1, j} \mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1}.$$

**3.3.1. Upper bound on  $\|\mathbf{V}_{k+1, j}\|_\infty$ .** The condition  $\|\mathbf{V}_{k+1, j}\|_\infty \leq 1$  can be dealt with using the same approach as before. We have that

$$V_{k+1, j}(s, t) = \tau \frac{\bar{\mathbf{R}}_{k+1, j}(s, :) \mathbf{v}_j}{\|\mathbf{v}_j\|_1}.$$

Since  $\tau \leq \tau_u$ , we can conclude from (3.121) that  $\tau \leq 0.9/(\mu_{ij} \sqrt{\phi_i \phi_j})$  for all  $i = k+1, \dots, k_0$ . Thus, we have that

$$\mathbb{P} \left( \tau \left| \frac{\bar{\mathbf{R}}_{k+1, j}(s, :) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > 1 \right) \leq \mathbb{P} \left( \tau \left| \frac{(\bar{\mathbf{R}}_{k+1, j}(s, :) - \mu_{i(s), j} \sqrt{\phi_{i(s)} \phi_j} \mathbf{e}_{n_j}^T) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > 0.1 \right),$$

where  $i(s)$  is the corresponding original block (row) index for the  $s$ th row of  $\bar{\mathbf{R}}_{k+1, j}$ . Since  $\|\mathbf{v}_j\| = 1$ ,  $(\bar{\mathbf{R}}_{k+1, j}(s, :)/\sqrt{\phi_{i(s)} \phi_j} - \mu_{i(s), j} \mathbf{e}_{n_j}^T) \mathbf{v}_j$  is a  $b$ -subgaussian random variable. Thus, by (3.2), we have that

$$\mathbb{P} \left( \tau \left| \frac{(\bar{\mathbf{R}}_{k+1, j}(s, :) - \mu_{i(s), j} \sqrt{\phi_{i(s)} \phi_j} \mathbf{e}_{n_j}^T) \mathbf{v}_j}{\|\mathbf{v}_j\|_1} \right| > 0.1 \right) \leq 2 \exp \left( -\frac{0.1^2 \|\mathbf{v}_j\|_1^2}{2b^2 \tau^2 \phi_{i(s)} \phi_j} \right).$$

To show this is exponentially small, we first analyze the denominator. We start by noting that

$$\begin{aligned} \tau^2 \phi_{i(s)} \phi_j &\leq (\tau'_u)^2 \phi_{i(s)} \phi_j \\ &\leq c_5(\mathbf{p})^2 \phi_{i(s)} \phi_j / \phi_1^2 \\ &\leq c_5(\mathbf{p})^2 c_0 \phi_j / \phi_1 \\ &\leq c_5(\mathbf{p})^2 c_0 \rho_\sigma (\rho_m \rho_n)^{1/2} \\ &\equiv c_8(\mathbf{p}). \end{aligned}$$

The second line was obtained from (3.117) and the third from (3.13), and the last line introduces another constant. Combining with (3.89) for the numerator, we obtain the following tail bound:

$$(3.100) \quad \mathbb{P} (\|\mathbf{V}_{k+1, j}\|_\infty > 1) \leq 2 \exp \left( -\frac{0.1^2 0.3^2 \delta_v^2 n_j}{2b^2 c_8(\mathbf{p})} \right).$$

**3.3.2. Upper bound on  $\|\mathbf{W}_{k+1, j}\|$ .** Now consider  $\mathbf{W}_{k+1, j}$ . It is clear that  $\mathbf{W}_{k+1, j} \mathbf{v}_j = \mathbf{0}$ . In addition, we have that

$$(3.101) \quad \mathbf{W}_{k+1, j} = \tau \theta \Phi \bar{\mathbf{Q}}_{k+1, j} \left( \mathbf{I} - \frac{\mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} \right),$$



where  $\bar{\mathbf{Q}}_{k+1,j} \in \mathbb{R}^{\bar{m}_{k+1} \times n_j}$  is a  $b$ -subgaussian matrix that is a concatenation of  $\mathbf{Q}_{lj}$ ,  $l = k+1, \dots, k_0$ , and

$$\Phi = \begin{pmatrix} \sqrt{\phi_{k+1}\phi_j} \mathbf{I}_{m_{k+1}} & & \\ & \ddots & \\ & & \sqrt{\phi_{k_0}\phi_j} \mathbf{I}_{m_{k_0}} \end{pmatrix}.$$

By the same argument as before,

$$(3.102) \quad \left\| \mathbf{I} - \frac{\mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} \right\| \leq 1 + \frac{1}{0.3\delta_{v,j}} \leq c_7(\mathbf{p}) + 1/2,$$

where  $c_7(\mathbf{p})$  was defined by (3.99). Also,

$$(3.103) \quad \|\Phi\| = \sqrt{\bar{\phi}_{k+1}\phi_j},$$

where  $\bar{\phi}_{k+1} = \max_{i=k+1, \dots, k_0} \phi_i$ . Now suppose

$$(3.104) \quad \|\bar{\mathbf{Q}}_{k+1,j}\| \leq c_9(\mathbf{p})(m_j n_j)^{1/2}/\sqrt{c_0},$$

where

$$(3.105) \quad c_9(\mathbf{p}) = \frac{0.47}{\rho_\sigma(c_7(\mathbf{p}) + 1/2)(k+1)},$$

and  $c_0$  was defined in (3.13). (Below we will argue that (3.104) happens with high probability.)

Using the hypothesis (3.104),

$$\begin{aligned} \|\mathbf{W}_{k+1,j}\| &\leq \tau\theta \|\Phi\| \cdot \left\| \mathbf{I} - \frac{\mathbf{v}_j \mathbf{e}_{n_j}^T}{\|\mathbf{v}_j\|_1} \right\| \cdot \|\bar{\mathbf{Q}}_{k+1,j}\| \\ &\leq \tau\theta \sqrt{\bar{\phi}_{k+1}\phi_j} (c_7(\mathbf{p}) + 1/2) c_9(\mathbf{p})(m_j n_j)^{1/2}/\sqrt{c_0} \\ &\leq \tau\theta \phi_j (c_7(\mathbf{p}) + 1/2) c_9(\mathbf{p})(m_j n_j)^{1/2} \\ &= \tau\theta \bar{\sigma}_j \frac{0.47}{\rho_\sigma(k+1)} \\ &\leq \frac{0.47\tau\theta}{k+1} \bar{\sigma}_k \\ &\leq \frac{1}{k+1} \cdot \min_{i=1, \dots, k} \sigma_k. \end{aligned}$$

The first line follows from (3.101), the second from (3.103), (3.102), and (3.104). The third and fifth follow from (3.13) and (3.11), respectively, and the last from (3.61).

Now we show that the hypothesis (3.104) holds with high probability using Lemma 3.2. As mentioned above,  $\bar{m}_{k+1}$  denotes the number of rows of  $\bar{\mathbf{Q}}_{k+1,j}$ , i.e.,  $m_{k+1} + \dots + m_{k_0}$ .

$$\begin{aligned} \mathbb{P} \left( \|\bar{\mathbf{Q}}_{k+1,j}\| > c_9(\mathbf{p})(m_j n_j)^{1/2}/\sqrt{c_0} \right) &\leq \exp \left( -\frac{8c_9(\mathbf{p})^2}{81b^2c_0} m_j n_j + (\log 7)(\bar{m}_{k+1} + n_j) \right) \\ &= \exp \left( -\frac{4c_9(\mathbf{p})^2}{81b^2c_0} m_j n_j + (\log 7)\bar{m}_{k+1} \right) \\ &\quad \cdot \exp \left( -\frac{4c_9(\mathbf{p})^2c_0}{81b^2} m_j n_j + (\log 7)n_j \right). \end{aligned}$$

The second exponent in the second line tends to  $-\infty$  linearly with  $m_j$ ; the first exponent also tends to  $-\infty$  linearly provided that

$$(3.106) \quad \frac{\bar{m}_{k+1}}{m_j n_j} \leq K < \frac{4c_9(\mathbf{p})^2}{81b^2 c_0 (\log 7)},$$

where  $K$  is some constant (independent of  $m_i, n_i$  for any  $i$ ), which holds under the assumption (3.14).

The analysis of the  $(i, k+1)$  block is similar for  $i = 1, \dots, k$ .

### 3.4. Analysis for block $(k+1, k+1)$ .

**3.4.1. Upper bound on  $\|\mathbf{V}_{k+1, k+1}\|_\infty$ .** For the last block  $(k+1, k+1)$ , we will simply construct  $\mathbf{V}_{k+1, k+1} \in \mathbb{R}^{\bar{m}_{k+1} \times \bar{n}_{k+1}}$  from  $(k_0 - k)^2$  subblocks  $\mathbf{V}_{st}^{(k+1)} \in \mathbb{R}^{m_s \times n_t}$ ,

$$\mathbf{V}_{st}^{(k+1)} = \tau \mu_{st} \sqrt{\phi_s \phi_t} \mathbf{e}_{m_s} \mathbf{e}_{n_t}^T, \quad s, t = k+1, \dots, k_0.$$

Since  $\tau \leq \tau_u$ , by (3.122) we have that  $\tau \leq 0.9/(\mu_{st} \sqrt{\phi_s \phi_t})$  for all  $s, t = k+1, \dots, k_0$ . Thus,  $\|\mathbf{V}_{k+1, k+1}\|_\infty \leq 1$ .

**3.4.2. Upper bound on  $\|\mathbf{W}_{k+1, k+1}\|$ .**  $\mathbf{W}_{k+1, k+1}$  is composed of blocks:  $\mathbf{W}_{s,t}^{(k+1)} = \tau \theta (\sqrt{\phi_s \phi_t} \bar{\mathbf{Q}}_{st} + \bar{\mathbf{B}}_{s,t})$ , where  $\bar{\mathbf{Q}}_{s,t} \in \mathbb{R}^{m_s \times n_t}$ . We will write the sum as

$$\mathbf{W}_{k+1, k+1} = \tau \theta (\Phi_2 \bar{\mathbf{Q}}_{k+1, k+1} \Phi_3 + \bar{\mathbf{B}}_{k+1, k+1}),$$

where  $\bar{\mathbf{Q}}_{k+1, k+1}$  contains entries chosen from a  $b$ -subgaussian distribution, and

$$\Phi_2 = \begin{pmatrix} \sqrt{\phi_{k+1}} \mathbf{I}_{m_{k+1}} & & \\ & \ddots & \\ & & \sqrt{\phi_{k_0}} \mathbf{I}_{m_{k_0}} \end{pmatrix}$$

and

$$\Phi_3 = \begin{pmatrix} \sqrt{\phi_{k+1}} \mathbf{I}_{n_{k+1}} & & \\ & \ddots & \\ & & \sqrt{\phi_{k_0}} \mathbf{I}_{n_{k_0}} \end{pmatrix}.$$

Clearly,  $\|\bar{\mathbf{B}}_{k+1, k+1}\| = \max_{l=k+1, \dots, k_0} \bar{\sigma}_l = \bar{\sigma}_{k+1}$  and  $\|\Phi_2\| = \|\Phi_3\| = (\bar{\phi}_{k+1})^{1/2}$ , where  $\bar{\phi}_{k+1}$  was defined as in (3.103). Thus

$$(3.107) \quad \|\mathbf{W}_{k+1, k+1}\| \leq \tau \theta \bar{\phi}_{k+1} \|\bar{\mathbf{Q}}_{k+1, k+1}\| + \tau \theta \bar{\sigma}_{k+1}.$$

Applying the assumption (3.12) to the second term of (3.107), we have that

$$\begin{aligned} \tau \theta \bar{\sigma}_{k+1} &\leq \frac{0.47 \tau \theta \bar{\sigma}_k}{2(k+1)} \\ &\leq \frac{1}{2(k+1)} \cdot \min_{i=1, \dots, k} \sigma_i. \end{aligned}$$

The second line follows from (3.61).

Now turning to the first term, let us suppose that

$$(3.108) \quad \|\bar{\mathbf{Q}}_{k+1,k+1}\| \leq \frac{0.23\sqrt{m_l n_l}}{(k+1)c_0},$$

where  $c_0$  is from (3.13) and  $l$  is the index of the min of  $\sigma_1, \dots, \sigma_k$ . (Below we will argue that this holds with probability exponentially close to 1.) Then

$$\begin{aligned} \tau\theta\bar{\phi}_{k+1} \|\bar{\mathbf{Q}}_{k+1,k+1}\| &\leq \frac{0.23\tau\theta\phi_{k+1}\sqrt{m_l n_l}}{(k+1)c_0} \\ &\leq \frac{0.23\tau\theta\phi_l\sqrt{m_l n_l}}{k+1} \\ &= \frac{0.46\tau\theta\bar{\sigma}_l}{2(k+1)} \\ &\leq \frac{1}{2(k+1)} \cdot \min_{i=1,\dots,k} \sigma_i. \end{aligned}$$

The second line uses (3.108) and the last line uses (3.61) and the choice of  $l$ . We have analyzed both terms of (3.107) and established  $\|\mathbf{W}_{k+1,k+1}\| \leq \frac{1}{k+1} \min_{i=1,\dots,k} \sigma_i$  as required.

We now analyze the probability that (3.108) fails. According to Lemma 3.2

$$\mathbb{P}\left(\|\bar{\mathbf{Q}}_{k+1,k+1}\| > \frac{0.23\sqrt{m_l n_l}}{(k+1)c_0}\right) \leq \exp\left(-\frac{8 \cdot 0.23^2 m_l n_l}{81b^2(k+1)c_0} + (\log 7)(\bar{m}_{k+1} + \bar{n}_{k+1})\right).$$

This quantity tends to zero exponentially fast as long as

$$(3.109) \quad \frac{\bar{m}_{k+1} + \bar{n}_{k+1}}{\min_{i=1,\dots,k} m_i n_i} \leq K < \frac{8 \cdot 0.23^2}{81b^2(k+1)c_0(\log 7)},$$

where  $K$  is some constant (independent of the matrix size), which holds under the assumption (3.14).

**3.5. Definitions of the scalars.** The definitions of the scalars appearing in the theorem and the proof can now be provided based on the inequalities developed during the proof.

We start by defining  $\tau_\ell$  as follows:

$$(3.110) \quad \tau_\ell = c_4(\mathbf{p})\phi_1^{-1},$$

where  $c_4(\mathbf{p})$  is defined as

$$(3.111) \quad c_4(\mathbf{p}) = \rho_\sigma \sqrt{\rho_m \rho_n} \max\left\{6(k+1)\sqrt{\rho_m \rho_n}, \frac{4}{\xi_u}, \frac{4}{\xi_v}, 2 + \frac{2}{\delta_u}, 2 + \frac{2}{\delta_v}\right\}.$$

Applying inequality (3.83), the following inequalities that have already been used in the preceding analysis indeed hold:

$$(3.112) \quad \tau_\ell \geq 6(k+1)\sqrt{\rho_m \rho_n} \max_{i=1,\dots,k} \phi_i^{-1},$$

$$(3.113) \quad \tau_\ell \geq \max\left\{\frac{4}{\xi_u}, \frac{4}{\xi_v}\right\} \max_{i=1,\dots,k} \phi_i^{-1},$$

$$(3.114) \quad \tau_\ell \geq \left(2 + \max\left\{\frac{2}{\delta_u}, \frac{2}{\delta_v}\right\}\right) \max_{i=1,\dots,k} \phi_i^{-1}.$$

The constant  $c_3(\mathbf{p})$  is then defined as follows:

$$(3.115) \quad c_3(\mathbf{p}) = \frac{1}{2} \left( 1.2^4 (c_4(\mathbf{p}))^2 \left( \frac{k\rho_m\rho_n}{k + \rho_m\rho_n - 1} \right) + 1 \right)^{-1/2}.$$

Next, we define

$$(3.116) \quad \tau_u = c_5(\mathbf{p})\phi_1^{-1},$$

where

$$(3.117) \quad c_5(\mathbf{p}) = \frac{4}{3} \left( \frac{1+(k-1)\sqrt{\rho_m\rho_n}}{1+(k-1)\rho_\sigma^{-2}} \right) + \frac{4}{3} \left( \sqrt{\left( \frac{1+(k-1)\sqrt{\rho_m\rho_n}}{1+(k-1)\rho_\sigma^{-2}} \right)^2 + \frac{(1+(k-1)\rho_m\rho_n)(c_3(\mathbf{p})^{-2}-1)}{1+(k-1)\rho_\sigma^{-2}}} \right).$$

Note that  $c_4(\mathbf{p}) = \frac{25}{36} \cdot \sqrt{\frac{k+\rho_m\rho_n-1}{k\rho_m\rho_n}} \cdot \sqrt{c_3(\mathbf{p})^{-2}/4-1}$  from (3.115), which implies  $c_4(\mathbf{p}) < c_5(\mathbf{p})$  or  $\tau_\ell < \tau_u$ .

We now define

$$(3.118) \quad c_2(\mathbf{p}, c_0) = \frac{1}{c_5(\mathbf{p})} \cdot \min \left\{ \frac{0.3}{\rho_\sigma \sqrt{\rho_m\rho_n}}, \frac{0.9}{(c_0\rho_\sigma \sqrt{\rho_m\rho_n})^{1/2}}, \frac{0.9}{\sqrt{c_0}} \right\}.$$

Clearly, since  $c_5(\mathbf{p}) \geq (4/3)(1 + (c_3(\mathbf{p}))^{-1}) \geq 4$  and  $\rho_\sigma, \rho_m, \rho_n \geq 1$ , we have that  $c_2(\mathbf{p}, c_0) \leq 0.075 < 0.08$ .

Now, using (3.83) and the upper bound  $\phi_i/\phi_j \leq c_0$  for all  $i = 1, \dots, k+1$  and all  $j = 1, \dots, k$  (a restatement of (3.13)), the following inequalities indeed hold:

$$(3.119) \quad \tau_u \leq \min_{i=1, \dots, k} \frac{0.3}{\mu_{ii}\phi_i},$$

$$(3.120) \quad \tau_u \leq \min_{i,j=1, \dots, k} \frac{0.3}{\mu_{ij}\sqrt{\phi_i\phi_j}},$$

$$(3.121) \quad \tau_u \leq \min_{i=k+1, \dots, k_0; j=1, \dots, k} \frac{0.9}{\max\{\mu_{ij}, \mu_{ji}\}\sqrt{\phi_i\phi_j}},$$

$$(3.122) \quad \tau_u \leq \min_{i,j=k+1, \dots, k_0} \frac{0.9}{\mu_{ij}\sqrt{\phi_i\phi_j}}.$$

The last scalar to define is  $c_1(\mathbf{p}, c_0, b)$ . We define it as follows:

$$(3.123) \quad c_1(\mathbf{p}, c_0, b) = \min \left( \frac{4c_9(\mathbf{p})^2}{81b^2c_0(\log 7)}, \frac{8 \cdot 0.23^2}{81b^2(k+1)c_0(\log 7)} \right),$$

where  $c_9(\mathbf{p})$  was defined by (3.105).

#### 4. Numerical examples.

**4.1. Biclique example.** We consider a simple example that involves a bipartite graph  $G = (U, V, E)$  with two nonoverlapping bicliques given by  $U_1 \times V_1$  and  $U_2 \times V_2$ , where  $U_1 \cap U_2 = \emptyset$  and  $V_1 \cap V_2 = \emptyset$ . The remaining edges in  $E$  are inserted at random with probability  $p$ . The  $U$ -to- $V$  adjacency matrix can be written in the form  $\mathbf{A} = \mathbf{B} + \mathbf{R}$ , where  $\mathbf{B}$  is a block diagonal matrix with  $k_0 = 3$  diagonal blocks, the

last of which is a block of all zeros, while the other two are blocks of all ones. If  $U_1 \cup U_2 = U$  and  $V_1 \cup V_2 = V$ , we can consider  $\mathbf{B}$  with just  $k_0 = 2$  diagonal blocks. We also assume that  $|U_1| = |U_2| = 1/2|U| = m/2$  and  $|V_1| = |V_2| = 1/2|V| = n/2$ . We would like to find these  $k = 2$  planted bicliques within the graph  $G$  under the presence of random noise simultaneously.

For this example,  $\bar{\mathbf{u}}_i = \mathbf{e}_{m_i}/\sqrt{m_i}$  and  $\bar{\mathbf{v}}_i = \mathbf{e}_{n_i}/\sqrt{n_i}$  for  $i = 1, 2$ . In addition,  $\bar{\sigma}_i = \sqrt{m_i n_i}$ ,  $i = 1, 2$ , which means  $\phi_1 = \phi_2 = 1$ . We can then choose  $\rho_u = \rho_v = 1$ ,  $\xi_u = \xi_v = 1$ ,  $\pi_u = \pi_v = 1$ ,  $\rho_m = \rho_n = 1$ , and  $\rho_\sigma = 1$ . Under the random setting described above,  $\mu_{ij} = p$  for all  $i \neq j = 1, 2$ . Given that  $k = k_0$ , we can set  $c_0 = 0$  and there is no need to consider the conditions related to noise blocks. With  $\bar{\mathbf{u}}_i = \mathbf{e}_{m_i}/\sqrt{m_i}$  and  $\bar{\mathbf{v}}_i = \mathbf{e}_{n_i}/\sqrt{n_i}$  for  $i = 1, 2$ , the analysis is simpler and we only need  $c_4(\mathbf{p}) = 2$  since (3.112) and (3.113) are not needed while (3.114) can be relaxed to  $\tau_\ell \geq 2 \max_{i=1, \dots, k} \phi_i^{-1}$ . The constant  $c_3(\mathbf{p})$  has a better approximation:

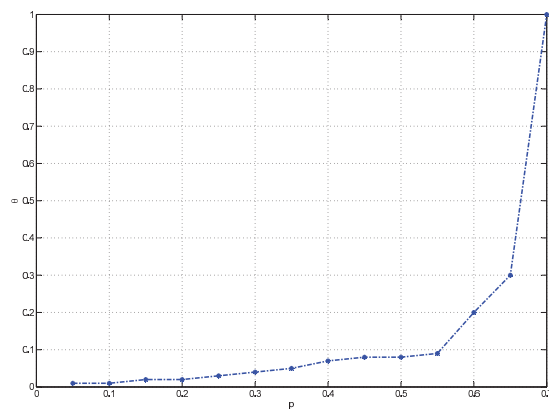
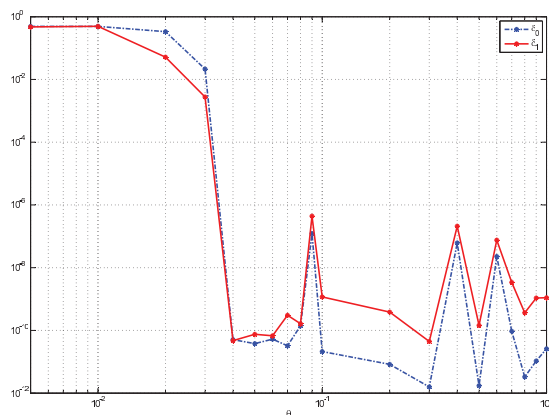
$$c_3(\mathbf{p}) = \frac{1}{2} \left( \frac{36}{25} c_4(\mathbf{p}) - 1 \right)^{-1} = \frac{25}{94} \approx 0.266.$$

We then can compute  $c_2(\mathbf{p}, c_0)$  as

$$c_2(\mathbf{p}, c_0) = 0.3/c_5(\mathbf{p}) = (0.9/4) \left( 1 + (c_3(\mathbf{p}))^{-1} \right)^{-1} \approx 0.047,$$

which means that with  $p \leq 0.047$ , we are able to recover two planted cliques using the proposed convex formulation in (2.8) with  $0.376 \cdot (mn)^{-1/2} \leq \theta \leq 0.752 \cdot (mn)^{-1/2}$  with high probability. The results are quite restricted given the way we construct the dual solutions solely based on matrices of all ones. Having said that, these conditions are theoretical sufficient conditions. Practically, the convex formulation (2.8) with a wider range of  $\theta$  can recover planted bicliques under the presence of more random noise, i.e., higher probability  $p$ . The numerical computation is performed with CVX [10] for the biclique example discussed here with  $m = n = 50$ . We test the problem with 10 values of  $p$  ranging from 0.05 to 0.95. For each value of  $p$ , we construct a random matrix  $\mathbf{A}$  and solve (2.8) with 20 different values of  $\theta$  ranging from 0.005 to 1.0. The solution  $\mathbf{X}$  is scaled so that the maximum value of its entries is 1. We compare  $\mathbf{X}$  and  $\mathbf{B}$  by taking the maximum differences between their entries in diagonal blocks of  $\mathbf{B}$ ,  $\delta_1$  and that of off-diagonal blocks,  $\delta_0$ . For this example, we are not able to recover two planted bicliques, i.e., the block diagonal structure of the matrix  $\mathbf{B}$ , with  $\theta = 0.005$  for any  $p$  given large values for  $\delta_0$  and  $\delta_1$ . It is due to the fact that for smaller values of  $\theta$ , the objective of achieving better rank-2 approximation is more prominent than the objective of achieving the sparse structure. In addition, we cannot recover the two bicliques for  $p \geq 0.75$ . Figure 1 shows the minimum values  $\theta_{\min}(p)$  of  $\theta$  with which (2.8) can be used to recover the planted bicliques when there is a significant reduction in the values of  $\delta_0$  and  $\delta_1$ . The graph indicates that we need larger  $\theta$  for the settings with more random noise. Figure 2 plots these differences (in log scale) for  $p = 0.30$  and we can see that  $\delta_0$  and  $\delta_1$  change from  $10^{-2}$  to  $10^{-10}$  between  $\theta = 0.03$  and  $\theta = 0.04$ . When the planted bicliques can be recovered, all of these values are in the order of  $10^{-6}$  or less, which indicates the recovery ability of our proposed formulation for this example under the presence of noise. Note that for this special example of binary data, the range of the values of  $\theta$  with which two planted bicliques can be recovered is usually large enough to cover the whole remaining interval  $[\theta_{\min}(p), 1]$  considered in this experiment.

Under the setting of this experiment, two blocks have the same size, i.e.,  $m_1 n_1 = m_2 n_2 = mn/4$ , which means  $\bar{\sigma}_1 = \bar{\sigma}_2$ . As mentioned previously, if we replace the Ky

FIG. 1. Minimum values of  $\theta$  to recover two planted bicliques for different values of  $p$ .FIG. 2. Maximum differences between entries in diagonal blocks and off-diagonal blocks for  $p = 0.30$ .

Fan 2- $k$ -norm in (2.8) by the Ky Fan  $k$ -norm, it is likely that we can still retrieve the information of singular vectors, which is enough for this experiment. We now run the Ky Fan  $k$ -norm formulation with different levels of noise by varying  $p$  from 0.05 to 0.95. Similarly, we also test the trace norm formulation proposed by Ames [1] under the Bernoulli model with  $\alpha = 1$  and  $\beta = p$  given this is a biclique instance. Figure 3 show the plots of  $\max\{\delta_0, \delta_1\}$  obtained from the three different models. It shows that all three models can handle noisy instances with  $p \leq 0.7$  with the trace norm model achieving the best result in terms of accuracy. It is due to the fact that if the trace norm model is successful, it returns the (unique) exact solution. In the next examples, we will demonstrate that if singular values are needed as parts of the recovery result, both the Ky Fan  $k$ -norm and the trace norm model are not able to deliver.

**4.2. Examples with synthetic gene expression data.** In this section, we apply our formulation for synthetic gene expression data sets studied in Prelić et al. [18]. Under this setting, biclusters are *transcription modules*, which are defined by a set of genes  $\mathcal{G}_i$  and a set of experimental conditions  $\mathcal{C}_i$ . Prelić et al. [18] provide two types of biclusters: constant clusters with binary gene expression matrices, which are similar to data inputs in the biclique problem, and additive clusters with integer

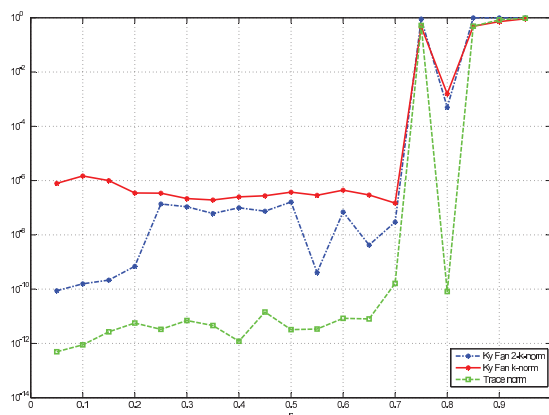


FIG. 3. Maximum differences between entries for three models.

gene expression matrices. We will focus on additive clusters in this section. Following Prelić et al. [18], we will examine the effects of noise with  $k = 10$  nonoverlapping transcription modules, each of which consists of 10 genes and 5 experimental conditions. The resulting gene expression matrices  $E$  are  $100 \times 50$  matrices with element values range from 0 to 100. Within the implanted biclusters, the values are at least 50, while the background values, i.e., outside the biclusters, are less than 50. Furthermore, average gene expression values are different from one implanted bicluster to another, and within each bicluster, the values are also different from one another. We add random normal noise,  $r_{ij} \sim N(0, (50\sigma)^2)$ , where  $\sigma$  is the noise level,  $0 \leq \sigma \leq 0.1$ , to the gene expression values while maintaining their nonnegativity, i.e.,  $e_{ij} \leftarrow \max\{e_{ij} + r_{ij}, 0\}$ . More details of how to construct these gene expression matrices can be found in Prelić et al. [18].

In order to compare different biclustering methods, Prelić et al. [18] defined a match score of two biclusters  $\mathcal{B} = (\mathcal{G}_i, \mathcal{C}_i)_{i=1,\dots,k}$  and  $\mathcal{B}' = (\mathcal{G}'_i, \mathcal{C}'_i)_{i=1,\dots,k}$  as

$$(4.1) \quad S_G^*(\mathcal{B}, \mathcal{B}') = \frac{1}{k} \sum_{i=1}^k \max_{j=1,\dots,k} \frac{|\mathcal{G}_i \cap \mathcal{G}'_j|}{|\mathcal{G}_i \cup \mathcal{G}'_j|}.$$

Clearly,  $S_G^*(\mathcal{B}, \mathcal{B}') \in [0, 1]$  and  $S_G^*(\mathcal{B}, \mathcal{B}') = 1$  if  $\mathcal{B}$  and  $\mathcal{B}'$  are the same. The match score is not symmetric, and given the implanted bicluster  $\mathcal{B}^*$ , each biclustering method with the resulting bicluster  $\mathcal{B}$  is measured by two measures, the *average bicluster relevance*,  $S_G^*(\mathcal{B}, \mathcal{B}^*)$  and the *average module recovery*,  $S_G^*(\mathcal{B}^*, \mathcal{B})$ . According to Prelić et al. [18], we can also define a similar match score  $S_C^*$  for experimental conditions. Having said that, to be consistent with the comparative study discussed in Prelić et al. [18], we will focus only on  $S_G^*$  match scores. In addition, for these gene expression applications, we also believe that it is of greater importance to correctly determine the clustering of the genes rather than of the experimental conditions. Now, for each noise level between 0 and 0.1, we will generate 10 noisy gene expression matrices, and as in Prelić et al. [18], the two performance measures will be averaged over these 10 instances. Similar to the biclique example, we solve (2.8) with 20 different values of  $\theta$  ranging from 0.005 to 1.0. For each run, the resulting matrix is scaled to best approximate the (noisy) input matrix, i.e., to minimize  $\|\alpha X^* - E\|$ , and element values are rounded down to zeros according to an appropriate threshold.



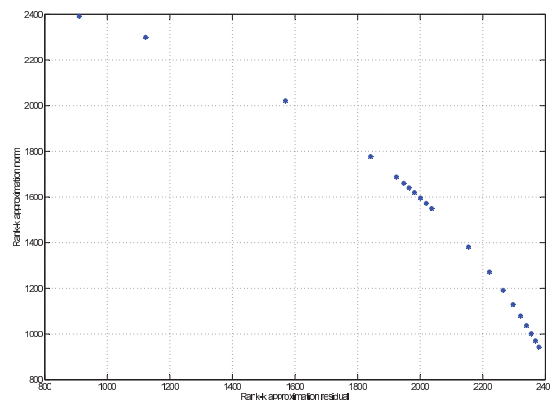


FIG. 4. Approximation averaging effect versus magnitude of resulting blocks.

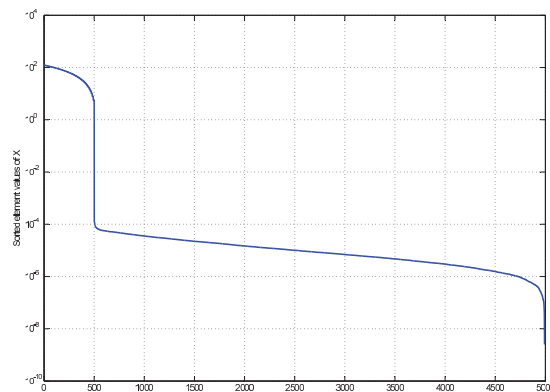
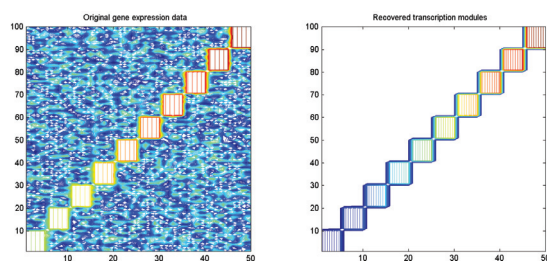
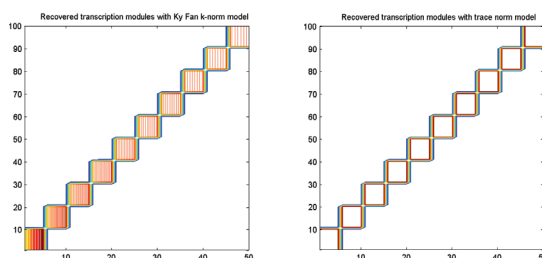
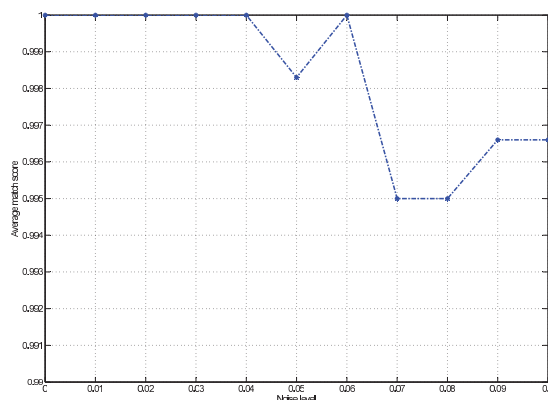


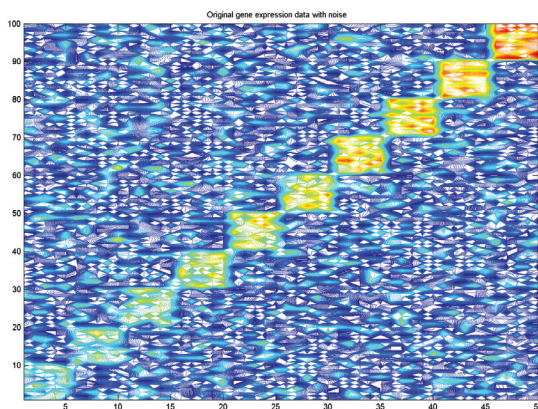
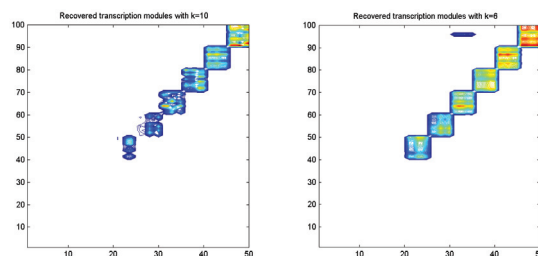
FIG. 5. Distinction between large and small element values of the resulting matrix.

The threshold is determined when there is a significant ratio (usually in the order of  $10^3$ ) between two consecutive sorted element values of the resulting matrix. The final computational issue is how to select the appropriate value for the parameter  $\theta$ . Theoretically, there is a range of  $\theta$  in which the recovery holds. For example, when all data blocks are square matrices of size  $n$ ,  $\theta$  is required to be in the order of  $O(1/(n\sqrt{k}))$ . Having said that, it is difficult to find correct constants in practice. For this particular example, we follow the heuristic used in Doan, Toh, and Vavasis [5], which finds the balance between the magnitude of the resulting matrix measured by the norm of its  $k$ -approximation and the approximation averaging effect measured by the norm of the residual. Figure 4 shows the plot of these two measures for our first run without noise ( $\sigma = 0$ ) and an appropriate value of  $\theta$  can be selected from the distinct middle range. We pick  $\theta = 0.07$ , which is in the middle of that range. Sorted element values of the resulting matrix are plotted in Figure 5 and we can see a significant transition (with a ratio of more than  $10^4$ ) between large and small values. The threshold for zero rounding can be set to be  $5 \times 10^{-4}$  in this case, knowing that all larger element values are larger than 5.

The recovered transcription modules are displayed in Figure 6 alongside the display of the original gene expression data. It clearly shows that all 10 transcription modules are recovered exactly, which means both performance measures, average bi-

FIG. 6. *Original gene expression data versus recovered transcription module.*FIG. 7. *Recovered transcription modules from two different models.*FIG. 8. *Match scores with different noise levels.*

cluster relevance and average module recovery, achieve the maximum value of 1. In addition, differences in gene expression levels between different implanted biclusters are present in the recovered transcription modules as in the original gene expression data. We also try to run the Ky Fan  $k$ -norm formulation and the trace norm model proposed by Ames [1] for the original gene expression data. Figure 7 shows the recovered transcription modules from the two models. Even though the recovered modules are correct, there is no significant difference in gene expression levels from one implanted bicluster to another as in the original gene expression data in the results of these two models. Furthermore, the trace norm model, which is developed for biclique problems, provides a single gene expression level within each implanted bicluster and this level is the same for all implanted biclusters. It shows that these two models cannot recover the information of singular values as expected.

FIG. 9. A noisy gene expression data matrix with  $\sigma = 0.3$ .FIG. 10. Recovery modules obtained with different  $k$ .

The effect of noise is captured in Figure 8. Both measures, average bicluster relevance and average module recovery, are the same in these instances and they are very close to 1 with the minimum value larger than 0.99. As compared to results reported in Prelić et al. [18, Figures 3(a), 3(b)], for this particular numerical example, our proposed method is comparable to (if not better than) the best algorithms such as BiMax, ISA, and Samba.

When the noise level goes higher, not all of the 10 modules can be recovered given the fact that the noisy background data can be misunderstood for actual expression data. Figure 9 shows an example of noisy gene expression data at the noise level of  $\sigma = 0.3$ . We run the proposed formulation with  $k = 10$  and recover 6 largest modules, which are not all perfect. We solve the problem again with  $k = 6$  instead and achieve much better results. The results are shown in Figure 10.

We conclude this section with a remark regarding algorithms used to solve the optimization problem (2.8). For the numerical examples discussed in this section, we solve its equivalent semidefinite optimization formulation (2.9) that involves semidefinite constraints for matrices of size  $(m + n) \times (m + n)$ . For instances with  $m = 50$  and  $n = 100$ , the computational time in 64-bit MATLAB 2013b with the CVX solver on our machine (3.50 GHz CPU and 16.0 GB RAM) is approximately 130 seconds. Clearly, for larger instances, we would need to develop appropriate first-order algorithms for the problem. A similar algorithmic framework as the one in Doan et al. [5] developed for the nuclear norm formulation could be an interesting topic for future research.

**5. Conclusions.** We have shown that a convex optimization problem with Ky Fan 2- $k$ -norm and  $\ell_1$ -norm can recover the  $k$  largest blocks of nonnegative block diagonal matrices under the presence of noise under certain conditions. This is an extension of the work in [6] and it could be used in biclustering applications.

**Acknowledgment.** We would like to thank two referees for their helpful comments and suggestions.

## REFERENCES

- [1] B. AMES, *Guaranteed clustering and biclustering via semidefinite programming*, Math. Program., 147 (2014), pp. 429–465.
- [2] A. ARGYRIOU, R. FOYGE, AND N. SREBRO, *Sparse prediction with the  $k$ -support norm*, in Advances in Neural Processing Systems 25, 2012, pp. 1466–1474.
- [3] R. BHATIA, *Matrix Analysis*, Grad. Texts in Math. 169, Springer-Verlag, New York, 1997.
- [4] I. DHILLON, *Co-clustering documents and words using bipartite spectral graph partitioning*, in Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'01), 2001, pp. 269–274.
- [5] X. V. DOAN, K.-C. TOH, AND S. VAVASIS, *A proximal point algorithm for sequential feature extraction applications*, SIAM J. Sci. Comput., 35 (2013), pp. A517–A540.
- [6] X. V. DOAN AND S. VAVASIS, *Finding approximately rank-one submatrices with the nuclear norm and  $\ell_1$ -norm*, SIAM J. Optim., 23 (2013), pp. 2502–2540.
- [7] N. FAN, N. BOYKO, AND P. PARDALOS, *Recent advances of data biclustering with application in computational neuroscience*, in Computational Neuroscience, W. Chaovalitwongse, P. Pardalos, and P. Xanthopoulos, eds., Springer, New York, 2010, pp. 85–112.
- [8] C. GIRAUD, *Low rank multivariate regression*, Electron. J. Stat., 5 (2011), pp. 775–799.
- [9] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, 1996.
- [10] M. GRANT AND S. BOYD, *CVX: MATLAB Software for Disciplined Convex Programming, Version 2.0 Beta*, <http://cvxr.com/cvx> (2013).
- [11] D. GROSS, *Recovering low-rank matrices from few coefficients in any basis*, IEEE Trans. Inform. Theory, 57 (2011), pp. 1548–1566.
- [12] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1990.
- [13] L. JACOB, F. BACH, AND J. P. VERT, *Clustered multi-task learning: A convex formulation*, in Advances in Neural Processing Systems 21, 2009, pp. 745–752.
- [14] M. LAURENT AND F. RENDL, *Semidefinite programming and integer programming*, in Handbook on Discrete Optimization, K. Aardal, G. Nemhauser, and R. Weismantel, eds., Elsevier, Amsterdam, 2005, pp. 393–514.
- [15] A. S. LEWIS, *The convex analysis of unitarily invariant matrix functions*, J. Convex Anal., 2 (1995), pp. 173–183.
- [16] S. MADEIRA AND A. OLIVEIRA, *Biclustering algorithms for biological data analysis: A survey*, IEEE/ACM Trans. Comput. Bio. Bioinform., 1 (2004), pp. 24–45.
- [17] A. McDONALD, M. PONTIL, AND D. STAMOS, *New Perspectives on  $k$ -Support and Cluster Norms*, <http://arxiv.org/abs/1403.1481> (2014).
- [18] A. PRELIĆ, S. BLEULER, P. ZIMMERMANN, A. WILLE, P. BÜHLMANN, W. GRUISSEM, L. HENNIG, L. THIELE, AND E. ZITZLER, *A systematic comparison and evaluation of biclustering methods for gene expression data*, Bioinformatics, 22 (2006), pp. 1122–1129.
- [19] A. TANAY, R. SHARAN, AND R. SHAMIR, *Discovering statistically significant biclusters in gene expression data*, Bioinformatics, 18 (2002), pp. S136–S144.
- [20] V. TIKHOMIROV, *Principles of extremum and application to some problems of analysis*, Pliska Stud. Math. Bulgar., 12 (1998), pp. 227–234.
- [21] G. A. WATSON, *On matrix approximation problems with Ky Fan  $k$  norms*, Numer. Algorithms, 5 (1993), pp. 263–272.
- [22] K. ZIĘTAK, *Properties of linear approximations of matrices in the spectral norm*, Linear Algebra Appl., 183 (1993), pp. 41–60.